DOCUMENT RESUME

ED 429 526 IR 019 434

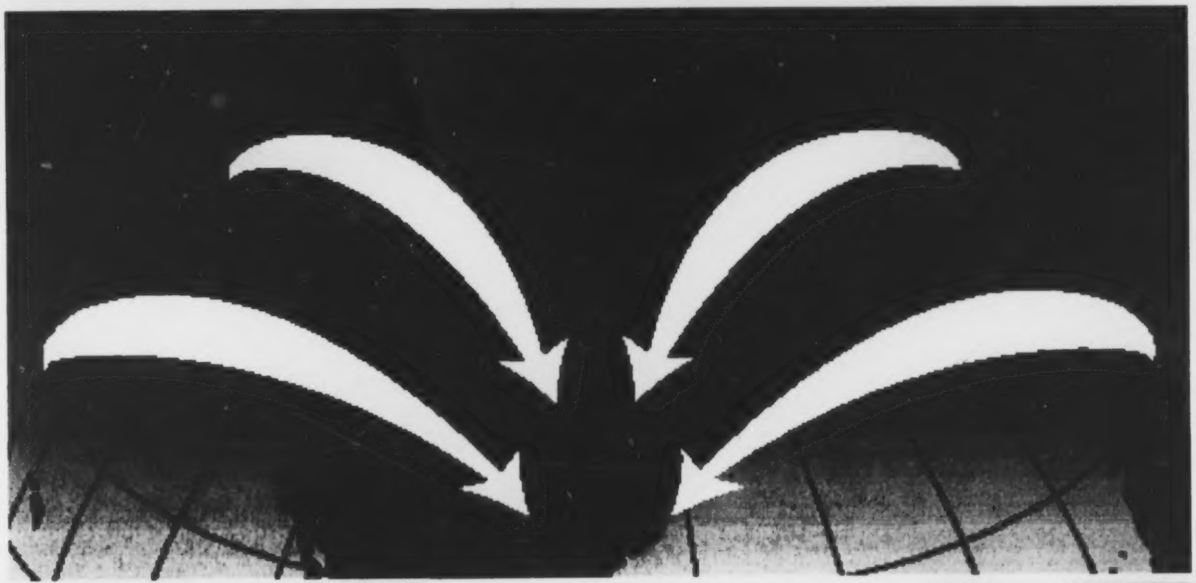
AUTHOR	Lobodzinski, Suave, Ed.; Tomek, Ivan, Ed.
TITLE	WebNet 97 World Conference of the WWW, Internet & Intranet Proceedings (2nd, Toronto, Canada, November 1-5, 1997).
ISBN	ISBN-1-880094-27-4
PUB DATE	1997-11-00
NOTE	1081p.; For individually selected papers, see IR 019 435-458. CD-ROM format only (includes proceedings from 1996, 1997, and 1998).
AVAILABLE FROM	Association for the Advancement of Computing in Education (AACE), P.O. Box 2966, Charlottesville, VA 22902; Web site: http://www.aace.org ($35 members; $40 non-members).
PUB TYPE	Collected Works - Proceedings (021)
EDRS PRICE	MF08/PC44 Plus Postage.
DESCRIPTORS	Computer Interfaces; *Computer Networks; Computer Security; Computer System Design; Computer Uses in Education; Conferences; Cooperative Learning; Courseware; Electronic Publishing; Hypermedia; Information Retrieval; Information Services; Information Systems; Legal Responsibility; Material Development; Multimedia Materials; Navigation (Information Systems); Privacy; Science and Society; Training; Virtual Reality; *World Wide Web
IDENTIFIERS	Browsing; *Intranets; Search Engines; Technology Implementation; Technology Utilization; Web Sites

ABSTRACT

The 1997 WebNet conference addressed research, new developments, and experiences related to the Internet and intranets. The 257 contributions of WebNet 97 contained in this proceedings comprise the full and short papers accepted for presentation at the conference. Included are positions papers by leading experts in the field; descriptions of ideas that are on the borderline between an idea, a prototype, and products; and reports on concrete applications of the World Wide Web, its impact on various aspects of life, and thoughts on how society might adjust to the resultant changes. Major areas covered include: advances in multimedia; browsing and navigation tools; computer-human interface issues; courseware development; educational multimedia on the Web; electronic publishing and the Web; industries and services; legal issues; psychology of Web use; security and privacy; statistical tools and user tracking; training; Web servers; application development tools; collaborative learning and work; country specific developments; data and link management; electronic commerce; future issues in WebNet technology; integration of Web applications and services; Internet-based multimedia/hypermedia systems; search engines; social and cultural issues; teaching; virtual reality; and Web site tools. (AEF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 429 526



WebNet 97
World Conference of the WWW, Internet, & Intranet

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY
G.H. Marks

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Edited by
Suave Lobodzinski
Ivan Tomek

**Proceedings of WebNet 97 —
World Conference of the WWW, Internet & Intranet**
Toronto, Canada; November 1-5, 1997

IR019434

AACE Association for the Advancement of Computing in Education

Copyright © 1997 by the Association for the Advancement of Computing in Education (AACE)

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the prior written permission of the publisher.

The publisher is not responsibile for the use which might be made of the information contained in this book.

Published by

Association for the Advancement of Computing in Education (AACE)
P.O. Box 2966
Charlottesville, VA 22902 USA

Printed in the USA

ISBN 1-880094-27-4

PREFACE

On behalf of the Program Committee and AACE, it is our pleasure to present to you the proceedings of the second WebNet conference—WebNet 97. Following in the footsteps of WebNet 96, this conference is addressing research, new developments, and experience related to the Internet and the Intranet.

Beginning with WebNet 96 held in San Francisco, WebNet conferences take place each year around late October to mid November. WebNet 97 is being held in Toronto, Canada , Oct. 30-Nov. 5, 1997, the venue of WebNet 98 will be Orlando, Florida; Nov. 7-12, 1998.

The 257 contributions of WebNet 97 presented in this volume are the Full and Short Papers accepted for presentation at the conference from a collection of more than 500 from 37 countries. All submissions were carefully reviewed by members of the Program Committee and their recommendations used for selection by the Program Chairs. The coverage of the contributions is very wide and this is one of the features that distinguishes WebNet from related conferences that focus on specific aspects of the Internet, World Wide Web, Hypertext, Multimedia, Global Networking, and related topics. Our intention is to provide an application oriented conference, a meeting place of developers, researchers and practitioners, with emphasis on the latter group and provide a forum where researchers, practitioners, and users from these disparate but related fields can meet and learn about new developments that impact their specialization.

As a consequence, this volume contains position papers by leading experts in the field; descriptions of ideas that are on the borderline between an idea, a prototype, and products; and reports on concrete applications of the Web; its impact on various aspects of life; and thoughts on how society will have to adjust to such changes and react to them. The areas covered at the conference and presented in this volume include:

Advances in Multimedia
Browsing and Navigation Tools
Computer-Human Interface (CHI) Issues
Courseware Development
Educational Multimedia on the Web
Electronic Publishing and the Web
Industries and Services
Legal Issues
Psychology of Web-Use
Security and Privacy
Statistical Tools and User Tracking
Training
Web Servers
Application Development Tools
Collaborative Learning and Work
Country Specific Developments
Data and Link Management
Electronic Commerce
Future Issues in WebNet Technology
Integration of Web Applications and Services
Net-Based Multimedia/Hypermedia Systems
Search Engines
Social and Cultural Issues
Teaching
Virtual Reality
Web Site Tools

In addition to the papers included in this volume, participants of this conference will also be able to listen to leading experts presenting Keynote and Invited lectures, and participate in tutorials, workshops, small-group discussions, panels, posters, and demonstrations. This printed record cannot show all aspects of a highly interactive, media-rich and Web-oriented meeting, but it does convey the depth and breath of the conference. It presents a snapshot of important and hot Web topics in the second half of 1997 and, with its 1997 predecessor and its successor WebNet volumes, it promises to become a milestone in the precipitous development of the Web.

Before you open the book and study the contributions, we wish you to enjoy this conference and this book...and to consider attending WebNet 97 or contributing to it. This is one of the best ways to stay current with the rapid and intriguing developments of the Web. Plan to periodically check **http://www.aace.org/conf/webnet** for the latest information!

In closing, we would like to thank all authors for submitting their work, and all members of the Program Committee listed on the following page for their cooperation and time spent reviewing the submissions. Special thanks go to Gary Marks (AACE) who is one of the main driving forces behind this volume and the WebNet conference, and his staff who did all the hard work required to get a large conference such as WebNet off the ground.

Ivan Tomek, Jodrey School of Computer Science, Acadia Univ., Canada
ivan.tomek@acadiau.ca
Suave Lobodzinski, California State Univ., USA
slobo@engr.csulb.edu
Chairs of WebNet 97 Program Committee

1997 Program Chairs

S.M. Lobodzinski; California State Univ. (USA)
Ivan Tomek; Acadia Univ. (Canada)

1997 Program Committee

Ron Baecker: Univ. of Toronto (Canada)
Philip Barker; Univ. of Teesside (UK)
Dines Bjoerner; UNU/IIST (Macao)
Meera Blattner; Lawrence Livermore Laboratory (USA)
Peter Brusilovsky; International Center of Science and Technology Information (Russia)
John Buford; Univ. of Massachusetts (USA)
Gordon Davies; Open Univ. (UK)
Roger Debreceny; Southern Cross Univ. (Australia)
Alain Derycke; Univ. de Lille (France)
Philip Doty; Univ. of Texas (USA)
Aude Dufresne; Univ. of Montreal (Canada)
Erik Duval; Katholieke Univ. Leuven (Belgium)
Adam Farquhar; Standord Univ. (USA)
Laura Fillmore; Online Bookstore (USA)
Richard Furuta; Texas A&M Univ. (USA)
Brian Gaines, Univ. of Calgary (Canada)
Colin Hensley; European Commission, Information Soc. Project Office (Belgium)
Rainer Kuhlen; Univ. of Konstanz (Germany)
John Leggett; Texas A&M; Univ. (USA)
Jennifer Lennon; Univ. of Auckland (New Zealand)
S.M. Lobodzinski; California State Univ. (USA)
Nadia Magnenat-Thalmann; Univ. of Geneva (Switzerland)
Gerald Maguire; Royal Technical Univ. (Sweden)
Fillia Makedon; Dartmouth College (USA)
Hermann Maurer; Graz Univ. of Technology (Austria)
Maria Teresa Molfino; Consiglio Nazionale delle Ricerche (Italy)
Max Muehlhaeuser; Univ. of Linz (Austria)
Douglas Norrie; Univ. of Calgary (Canada)
Andrew Odlyzko; AT&T; Bell Laboratories (USA)
George Olinger; A-Tek Solutions, USA
Henk Olivie; Katholieke Univ. Leuven (Belgium)
Charles Owen; Dartmouth College (USA)
Jacob Palme; Univ. of Stockholm (Sweden)
Gilbert Paquette; Tele-Univ. (Canada)
Elisabeth Patschke; Univ. of Freiburg (Germany)
Valery Petrushin; Georgia Tech Univ. (USA)
Reinhard Posch; Graz Univ.of Technology (Austria)
Raghu Ramakrishnan; Univ. of Wisconsin (USA)
Sam Rebelsky; Dartmouth College (USA)
Nick Scherbakov; Graz Univ. of Technology (Austria)
Wolfgang Schinagl; WIFI-IIC (Austria)
Gunter Schlageter; Univ. of Hagen (Germany)
John Schnase; Washington Univ. (USA)
Daniel K. Schneider; Univ. of Geneva (Switzerland)
Mildred Shaw; Univ. of Calgary (Canada)
Manolis Skordalakis; National Technical Univ. of Athens (Greece)
Klaus Tochtermann; Texas A&M Univ. (USA)
Ivan Tomek; Acadia Univ. (Canada)

Table of Contents

AWARD PAPERS

Implementing A Digital Signature Infrastructure for Web Applications
Lalitha Bhat & Chencheng Qiu, Bell Atlantic, USA
In Search of Web Forms: a Neural Agent for Organizing "Reusable" Notes Documents
Alberto Faro & Corrado Santoro, Universita' di Catania, Italy; Daniela Giordano, Concordia University, Canada
Beyond Java: An Infrastructure for High-Performance Mobile Code on the World Wide Web
Michael Franz, University of California, USA
Fetal Echocardiography Homepage
Krzysztof P. Wróblewski & Piotr M. Wróblewski, University of Pennsylvania, USA; Zhi Yun Tian & Jack Rychik, Children's Hospital of Philadelphia, USA

BUSINESS/CORPORATE SESSIONS

International Tele-Education Experiment using CAI System on the World-Wide Web
Yasuhisa Kato, Jun Hirai & Yoshimi Fukuhara, NTT Information and Communication Systems Laboratories, Japan; Mitsuru Yamada, KDD, Japan
Remote Lecture between USA and Japan over the ATM Network
Naoko Takahashi, Hisasumi Tsuchida & Yoshimi Fukuhara, NTT Information and Communication Systems Laboratories, Japan; Yoshiyasu Takefuji, KEIO University, Japan; Raymond K., Case Western Reserve University, USA
Working Together While Being S-e-p-a-r-a-t-e-d
G. van den Broek, Philips Research IST, Netherlands; F.J. Stewing, SNI C-lab, Germany

FULL PAPERS

Shared User Behavior on the World Wide Web
Ghaleb Abdulla, Edward A. Fox & Marc Abrams, Virginia Tech University, USA
Business Use of the Internet in New Zealand: A Follow-Up Study
Walt Abell, Lincoln University, New Zealand & Steve Black, Information Tools Ltd, New Zealand
Easy Ed: An Integration of Technologies for Multimedia Education
G. Ahanger & T.D.C. Little, Boston University, USA
A User-Adapted Interface for a Search Engine on the World Wide Web
Leonardo Ambrosini, Vincenzo Cirillo & Alessandro Micarelli, Università di Roma Tre, Italy
Techniques For Enhancing Web-Based Education
Kathy Barbieri & Susan Mehringer, Cornell University, USA
ClassACT: The Evolution of a Collaborative Learning Tool
Richard Barone & William Parod, Jr, Northwestern University, USA
Authoring and Development in an Online Environment: Web-based Instruction Using SGML
Paul Beam, The University of Waterloo, Canada
The Freedom to Choose: Transforming Content On-demand in the BSCW Shared Workspace System
Richard Bentley, Torsten Täschner & Wolfgang Appelt, German National Research Centre for Computer Science, Germany
Website Authoring using Dynamic Hypertext
Rick Bodner, Mark Chignell & Jim Tam, University of Toronto, Canada
How to Make Money with Web Access
Manfred Bogen & Guido Hansen, GMD - German National Research Center for Information Technology, Germany
Is the Web a Secure Environment for Electronic Commerce?
Manfred Bogen, Michael Lenz, Andreas Reichpietsch & Peter Simons, GMD - German National Research Center for Information Technology, Germany
Information Brokering on the World Wide Web
Stéphane Bressan & Thomas Lee, MIT Sloan School of Management, USA

Teams, Tasks, and Notices: Managing Collaboration via the World Wide Web
Charles L. Brooks, Frederick J. Hirsch & W. Scott Meeks, The Open Group Research Institute, USA
Developing and Integrating a Web-based Quiz into the Curriculum
Angela Carbone & Peter Schendzielorz, Monash University, Australia
Putting Usability First in the Design of Web Sites
Jim Carter, University of Saskatchewan, Canada
WebCiao: A Website Visualization and Tracking System
Yih-Farn Chen & Eleftherios Koutsofios, AT&T Labs - Research, USA
Comparative Diffusion of the Telephone and the World Wide Web: An Analysis of Rates of Adoption
Hsiang Chen & Kevin Crowston, Syracuse University, USA
Detecting Themes in Web Document Descriptors
David K.Y. Chiu & David R. Brooks, University of Guelph, Canada
Building a Configuration Assistant on the Web
Antonio Cisternino, Giuseppe Attardi & Maria Simi, Università di Pisa, Italy
WebSteps, a Tool for Synthesizing and Managing Web-based Business Processes
Jürgen Coppens, Oscar Figueiredo, Roger D. Hersch & Benoît Gennart, Ecole Polytechnique Fédérale, Switzerland; André-R. Probst & Maia Wentland Forte, University of Lausanne, Switzerland; Frédéric Vianin, Exel&Scopus, Switzerland
Web Host Enumeration Through DNS
Domenico Dato, Antonio Gullì & Giuseppe Attardi, Università di Pisa, Italy
Creating Adaptive Hyperdocuments for and on the Web
Paul De Bra, Eindhoven University of Technology, The Netherlands & University of Antwerp, Belgium; Licia Calvi, University of Antwerp, Belgium
Virtually Deschooling Society: Authentic Collaborative Learning via the Internet
R. T. Jim Eales & Laura M. Byrd, Virginia Tech, USA
Medtec: A Web-Based Intelligent Tutor for Basic Anatomy
Christopher R. Eliot, Daniel E. Neiman & Michelle M. LaMar, University of Massachusetts, USA
It's life Jim, but not as we know it!
Henrik Fagrell, Göteborg University, Sweden; Carsten Sørensen, Aston University, United Kingdom & Göteborg University, Sweden
'Agentizing' the Internet – the Practical Design of a Networked Mobile Agent System
Benjamin Falchuk & Ahmed Karmouch, University of Ottawa, Canada
Publishing Troubleshooting Expertise on the World-Wide Web: A Case Study
Robert Farrell & Frank T. Jones III, Bellcore, USA; Karen Kukich, Educational Testing Service, USA
Creating a Workshop on the World Wide Web: The Experiences of Internet 101
Carolyn Gale, John R. Bourne & Benjamin Thompson, Vanderbilt University, USA
Development and Integration of WWW-based Services in an Existing University Environment
John Garofalakis, Panagiotis Kappos, Athanasios Tsakalidis, John Tsaknakis, Giannis Tzimas & Vassilios Vassiliadis, University of Patras and Computer Technology Institute, Greece
Adaptive and Agile Interactive Learning Environments on the WWW
Sylvain Giroux, Cognicase, Canada; Richard Hotte & Kim Dao, Télé-Université, Canada
MiMi: A Java Implementation of the MicroMint Scheme
Vesna Hassler, Robert Bihlmeyer, Michael Fischer & Manfred Hauswirth, Technical University of Vienna, Austria
A Toolset for Personal Information Management
Daniel J. Helm & Raymond J. D'Amore, The MITRE Corporation, USA
World Wide Web Presentations for Volatile Hypermedia Database Output
Geert-Jan Houben, Eindhoven University of Technology, The Netherlands; Paul De Bra, Eindhoven University of Technology, The Netherlands & University of Antwerp, Belgium
Teaching Biology on the Internet
Thomas S. Ingebritsen, George G. Brown & John M. Pleasants, Iowa State University, USA
Information Access in the Web
Luca Iocchi & Daniele Nardi, Università di Roma, Italy
'Interactivity' – Tracking a New Concept
Jens F. Jensen, Aalborg University, Denmark

Disha: A Direction Toward Software Engineering on the Net
Raman Kannan, Zhen Chen, William Schroeder, Bo Cao, Rajesh Rao, Joe Occoner, Veenatejaswini Sanjivi, Kavitha Challapalli, Nanditha Gade & Bandish Chitre, Monmouth University, USA
Content is Key: Viewing Conceptualized Content
Robert E. Kent, Washington State University, USA; Christian Neuss, Technische Hochschule Darmstadt, Germany
On Developing a Web-Based Operating Environment
Lee Kennedy & Jim Carter, University of Saskatchewan, Canada
Cineast - An Extensible Web Browser
Eckhart Köppen & Gustaf Neumann, University of Essen, Germany; Stefan Nusser, Vienna University of Economics and Business Administration, Austria
Internet Based Collaboration: the Presence and Perspectives
Miloš Kravcík, Comenius University, Slovakia
A Framework for WWW-based Learning with Flexible Navigational Guidance
Christian Langenbach & Freimut Bodendorf, University of Erlangen-Nuremberg, Germany
Interface Design Issues for Web-based Collaborative Learning Systems
Glenn E. Lautenbacher & Dirk E. Mahling, University of Pittsburgh, USA
Experiences in Virtual Teaching
Karen A. Lemone, Worcester Polytechnic Institute, USA
A Taxonomy of Internet Communication Tools
Byron Long & Ronald Baecker, University of Toronto, Canada
Student Roles on the World Wide Web: Scholar, Researcher, Beggar, or Thief
Stephanie Ludi, Arizona State University, USA
Computer Mediated Communications: for Dabblers or Practitioners?
Jon Mason & Matthew Riddle, The University of Melbourne, Australia
Diversity in Online Courses
Robin Mason, The Open University, England
Cost Reduction through an Intranet: The Paperless Office
David Matusow, Joe Sparmo & John Bristow, NASA/Goddard Space Flight Center, USA
On Creating Hierarchical, Interlinked FAQ Archives
Tom Murphy, Silicon Graphics, USA
High Performance Web Server System: TCP Based Switching Approach
B. Nandy, K.L.E. Law & A.S.J. Chapman, Nortel Technology, Canada
Connection Caching for WWW Servers and Proxies
Noritaka Osawa, Kazuhisa Yanagawa & Toshitsuga Yuba, The University of Electro-Communications, Japan
ASML: Automatic Site Markup Language
Charles B. Owen, Fillia Makedon, Glen Frank & Michael Kenyon, Dartmouth Experimental Visualization Laboratory (DEVLAB), USA
Virtual Partnerships in Research and Education
Deborah A. Payne, Kelly A. Keating & James D. Myers, Pacific Northwest National Laboratory, USA
Telemedicine: An Inquiry in the Economic and Social Dynamics of Communications Technologies in the Medical Field
Francis Pereira, Elizabeth Fife & Antonio A. Schuh, University of Southern California, USA
Web-Based Educational Media: Issues and Empirical Test of Learning
Senthil Radhakrishnan & James E. Bailey, Arizona State University, USA
Visualization in a Mobile WWW Environment
Alberto B. Raposo & Ivan L. M. Ricarte, State University of Campinas (UNICAMP), Brazil; Luc Neumann, Computer Graphics Center (ZGDV), Germany; Léo P. Magalhães, University of Waterloo, Canada & State University of Campinas (UNICAMP), Brazil
Intelligent Control of Dynamic Caching Strategies for Web Servers and Clients
Mike Reddy & Graham P. Fletcher, University of Glamorgan, Wales, United Kingdom
A Database Architecture for Web-Based Distance Education
Daniel R. Rehak, Carnegie Mellon University, USA
Paper Interfaces to the World-Wide Web
Peter Robinson, Dan Sheppard, Richard Watts, Robert Harding & Steve Lay, University of Cambridge, England

Free Speech on the Internet: Legal, Social, and Political Issues
Richard S. Rosenberg, The University of British Columbia, Canada
Web-based Course Delivery and Administration using Scheme
Filippo A. Salustri, The University of Windsor, Canada
Law's Domain: Trademark Protection for Internet Address Names?
Kurt M. Saunders, Duquesne University, USA
WWW Tools for Accessing Botanical Collections
Erich R. Schneider, John J. Leggett, Richard K. Furuta, Hugh D. Wilson & Stephan L. Hatch, Texas A&M University, USA
The Internet in School: The Shaping of Use by Organizational, Structural, and Cultural Factors
Janet Ward Schofield & Ann Locke Davidson, University of Pittsburgh, USA
Legal Issues in Cyberspace
Alfred J. Sciarrino & Jack S. Cook, State University of New York at Geneseo, USA
The Intranet as a Cognitive Architecture for Training and Education: Basic Assumptions and Development Issues
Ahmed Seffah & Robert Maurice Bouchard, McGill University, Canada
An Architectural Framework for Developing Web-Based Interactive Applications
Ahmed Seffah, McGill College, Canada; Ramzan Ali Khuwaja, Nortel Technology, Canada
Teaching Cooperative Task Using the Web
Tadié Guepfu Serge & Frasson Claude, Université de Montréal, Canada; Lefebvre Bernard, Université du Québec à Montréal, Canada
A Foot In Both Camps: Interventions in National and Global Regulatory Processes by Nation-based Internet Organisations.
Jenny Shearer, University of Auckland, New Zealand
A Survey of Web Information Systems
Alberto Silva & José Delgado, IST/INESC; M. Mira da Silva, University of Evora
Design Considerations in Converting a Standup Training Class to Web-based Training: Some Guidelines from Cognitive Flexibility Theory
Nancee Simonson, The Bureau of National Affairs, Inc., USA
Authoring Tools for Courseware on WWW : the W3Lessonware Project
Phil Siviter, University of Brighton, UK
Cultural Impacts of the Internet and World Wide Web on a Computer-Literate Government Organization
Joe Sparmo, David Matusow & John Bristow, NASA/Goddard Space Flight Center, USA
Web-Based Requirements Analysis
Gees C. Stein, Karan Harbison & Stephen P. Hufnagel, University of Texas at Arlington, USA; James M. Mantock, ScenPro, Inc., USA
A 3D Topographic Map Viewer for the USA
S. Augustine Su, University of Maryland, USA; Shueh-Cheng Hu & Richard Furuta, Texas A&M University, USA
Automatic Interests Extraction Chasing the Browsing and Event History
Seiji Susaki & Tatsuya Muramoto, NTT Information and Communication Systems Labs, Japan
The Online Learning Academy
Suzanne Liebowitz Taylor & Donald P. McKay, Lockheed Martin C2 Integration Systems, USA; Ann Culp, Educational Technologies, USA; Stephen Baumann & Karen Elinich, The Franklin Institute Science Museum, USA
On the Use of Librarians Selection Routines in Web Search
Avgoustos. A. Tsinakos & Kostantinos. G. Margaritis, University of Macedonia, Greece
Modelling Alter-egos in Cyberspace Using a Work Flow Management Tool: Who Takes Care of Security and Privacy?
Reind P. van de Riet & Hans F.M. Burg, Vrije Universiteit, Amsterdam
Organising Distance Learning Process thanks to Asynchronous Structured Conversations
Claude Viéville & Frédéric Hoogstoël, Université des Sciences et Technologies de Lille, France
Incorporating Shared Relevance Feedback into a Web Search Engine
Kevin Y. Wei, Microstrategy, Inc., USA; Gregory C. Sharp, University of Wisconsin-Madison, USA

Cultural Implications in the World Wide Web: A Case for Research

Martyn Wild, Janice Burn, Ron Oliver & Sue Stoney, Edith Cowan University, Australia; Lyn Henderson, James Cook University, Australia

PANELS

Wired Communities - Can We Create Affinity Through Electronic Vicinity?: A PaneL

Ron Riesenbach, Telepresence Systems, Inc., Canada; Tom Jurenka, Disus, Canada; Gale Moore, University of Toronto, Canada

POSTER/DEMONSTRATIONS

A JAVA Toolkit for Representing a 3D Environment

Marco Aiello, Pierluca De Maria & Cristiano De Mei, Università di Roma La Sapienza, Italy

Utilizing Technology to Enable Practitioners to Serve as Mentors for Preservice Teachers: Strategies for Research

Gregory F. Aloia, Ramesh Chaudhari, Jeffrey P. Bakken & John Beemsterboer, Illinois State University, USA

Using Evaluation to Monitor and Improve the Early Implementation of Online University Teaching: an Australian Example

Pamela F. Andrew & John M. Owen, The University of Melbourne, Australia

Enabling Distance Education over the World Wide Web

I. Antoniou, University of Patras, Greece; C. Bouras, P. Lampsas & P. Spirakis, University of Patras & Computer Technology Institute, Greece

Growing Usage of Internet Services: A View From A Developing Country

Cengiz S. Askun & Kursat Cagiltay, Indiana University, USA

The World Wide Web: A new approach to world wide supply of epidemiological data on diabetes mellitus

Thomas Baehring, Hardy Schulze & Stefan R. Bornstein, University Hospital of Leipzig, Germany; Werner A. Scherbaum, University of Düsseldorf, Germany

Changing Ages: Transforming Paradigms, Policy andPedagogical Practice

Ivan W. Banks, Ruth R. Searcy, & Mike Omoregie, Jackson State University, USA

Real-Time Streaming Video for Ethernet-based Intranets

Jerzy A. Barchanski, Brock University, Canada

The Noteless Classroom

Robert Beard, Bucknell University, USA

Information Technology in the Second/Foreign Language Classroom: A Unit on Food

Mercè Bernaus, Departament d'Ensenyament, Spain

Large Scale Remote Graduate Instruction in Beam Physics

Martin Berz, Béla Erdélyi & Jens Hoefkens, Michigan State University, USA

Mass Media and Identity

Karin Blair, Switzerland

MusicWeb, New Communication and Information Technologies in the Music Classroom

Carola Boehm & Celia Duffy, University of Glasgow, United Kingdom

Dynamic Web Access for Collaborative Writing

Jennifer J. Burg, Anne Boyle, Yinghui Wu, Yue-Ling Wong & Ching-Wan Yip, Wake Forest University, USA

The National Ergonomical Information Network of Ukraine

Alexander Burov, National Research Institute for Design, Ukraine

Graphical Representation of Students' Laboratory Marks on the World Wide Web

Angela Carbone, Monash University, Australia

Selecting the Right Person For the Job - An Interactive Tutor Recruitment package

Angela Carbone, Monash University, Australia

Open Standard Content Cookies: Utility vs. Privacy

Chad Childers, Ford Motor Company, USA; Iain O'Cain, Intranet Org, Canada; Linda Bangert, Internet Education Group, USA; Mike O'Connor, Silicon Graphics, Inc., USA

A Framework for the Comparison of Computer-Supported Collaborative Learning Applications

Ruth Crawley & Jerome Leary, University of Brighton, United Kingdom

World Wide Web and Data Base Managers: Threads Woven into Fabric
Gregory M. Dick & J. Jeffrey Sernell, University of Pittsburgh at Johnstown, USA; Scott Koontz, S. E. Koontz Software Solutions, USA
Vocal Point: A Collaborative, Student Run On-Line Newspaper
Scott Dixon & Chrissy Anderson, Centennial Middle School, Colorado, USA
Utilizing the WWW for Industrial Training
Eman El-Sheikh, Fran Bakowska, Chris Penney, Rong Liu & Jon Sticklen, Michigan State University, USA
A Practical Approach and Infrastructure for Large Scale Web Applications and Services
Pierre France & Michele Kearney, Systems & Computer Technology Corporation, USA
Election Project Experience
Ari Frazão, Fabiola Greco, Lúcia Melo & Teresa Moura, Brazilian Research Netword (RNP), Brazil
Designing Web-Based Instruction for High School Courses
Jed Friedrichsen, Marilyn Altman & Cynthia Blodgett-McDeavitt, University of Nebraska-Lincoln, USA
Text Generation in Business Object Frameworks
M.O. Froehlich & R.P. van de Riet, Vrije Universiteit, The Netherlands
Pornography is Not The Problem: Student use of the Internet as an Information Source
Katherine J. Fugitt, University of Washington, USA
Empirical Analysis of the Use of Electronic Bulletin Boards Supplementing Face to Face Teaching
Jayne Gackenbach, Athabasca University, Canada
Solution Prototype for the Adaptation of an Information System into an Intranet/Internet Environment under Windows95
A. García-Crespo, P. Domingo, F. Paniagua, E. Jarab & B. Ruiz, Universidad Carlos III de Madrid, Spain
A New Twist to an Old Idea - Telementoring Using the Web
Melanie Goldman, BBN Corporation, USA
A Class Web Page Comes to Life
Patricia Gray, Rhodes College, USA
Using Faculty Focus Groups to Conceptualize a Case Seminar Facility for Distance Education Courses at The Medical College of Georgia
Kathleen M. Hannafin & Shary L. Karlin, Medical College of Georgia, USA
Integrating On-Line And Face-To-Face Work In Professional And Learning Environments
Don Harben, Toronto Board of Education, Canada; John Myers, OISE/UT, Canada
Educators in Instructional Technology
Jeffery L. Hart, Utah State University, USA
A Methodology for Determining Website Navigational Efficiency
Jeffrey B. Hecht, Illinois State University, USA; Perry L. Schoon, Florida Atlantic University, USA
The Rhythm of the Web: Patterns of "Multiple N's of One"
Andrew Henry & Valerie L. Worthington, Michigan State University, USA
Using Internet Tools to Create Cross-Disciplinar y, Collaborative Virtual Learning Environments
Peggy Hines, Phyllis Oakes, Calvin Lindell & Donna Corley, Morehead State University, USA
Collaborative diagnosing and distance-learning materials for medical professionals
Takahide Hoshide, Yasuhisa Kato, Yoshimi Fukuhara, Makoto Akaishi, David W. Piraino, James D. Thomas & Mitsuru Yamada, NTT Information and Communication Systems Laboratories, Japan
Using Web Sites in University Courses as Bulletin Boards and for Enrichment
M. Eleanor Irwin, University of Toronto at Scarborough, Canada
A path over the Internet to a student-centered on-line classroom: The SUNY Learning Network — the design, development of nineteen SLN Web-based courses
Mingming Jiang, State University of New York, USA
Windows to the Universe: An Internet-Based Educational Resource for the General Public
Roberta M. Johnson, The University of Michigan, USA
Use of Browser-based Technology in Undergraduate Medical Education Curriculum
Laleh S. Khonsari, University of South Florida, USA
Developing Internet in Belarus: Minsk Internet Project
Sergei Kritsky, Nikolay Listopad & Igor Tavgen, Belarussian State University of Informatics

Collaborative Teaching in Cyberspace
Thérèse Laferrière, Laval University, Canada
Multimedia Presentations in Life Sciences Teaching
Revital Lavy, Nurit Wengier, Booki Kimchi & Rina Ben-Yaacov, The Open University of Israel, Israel
Mentoring an Internet-based Distance Education Course: Problems, Pitfalls, and Solutions
Marcia L. Marcolini & LeeAnn Hill, West Virginia University, USA
The CREN Virtual Seminar Series: Learning at Your Desktop
Gregory A. Marks, Merit Network, Inc., USA; Susan Gardner, Gardner Communications, USA; Rick Witten, Synapsys Media Network, Inc., USA
A Web-based Course in English as a Second Language: A Case Study
Ian Marquis & Tài Nguyen, North York Board of Education, Canada; Jean Wang, Simon Fraser University, Canada
Benjamin Franklin House: An illustration of a site management and visual design tool for complex, multi-authored web sites
Gil E. Marsden, Gareth J. Palmer & Harold Thimbleby, Middlesex University, England
Distance Education Based on Computer Textbooks
Dmitry Sh. Matros, University of Pedagogic, Russia
Political Philosophy and the Technology Curriculum
Bruce W McMillan, University of Otago, New Zealand
What it Really Takes to Put Your Lab on the Internet
John Mertes, Rhodes School, USA
Web-based Education: Considerations and Approaches
Jay Moonah, Ryerson Polytechnic University, Canada
Multi Media - Malaysian National Curriculum
Vijaya Kumaran K.K. Nair, Stamford College Berhad, Malaysia
How to Provide Public Address to Internet Information Sources on Public Access Workstations?
Paul Nieuwenhuysen, Vrije Universiteit Brussel, Belgium
Global Educational Database on The WWW(World-Wide Web) and Its Application in School
Masatoyo Ohshima, Kanzaki-Seimei Senior High School, Japan; Yasuhisa Okazaki, Hisaharu Tanaka & Hiroki Kondo, Saga University, Japan; Hiroshi Nokita, Yamato Junior High School, Japan; Hidekatsu Hara, Saga Prefectural Government, Japan; Hirofumi Momii, Saga Prefectural Education Center, Japan; Kenzi Watanabe, Wakayama University, Japan
Canada's Wired Writers: The Writers In Electronic Residence Program
Trevor Owen, York University, Canada
SELENA: Walking on the Moon or How to Make Decisions Using the Web
Valery A. Petrushin, Mark Yohannan & Tetyana Lysyuk, Georgia Institute of Technology, USA
Issues of Authority in On-line Instruction
JoAnne Podis, Ursuline College, USA
Codes of Ethics for Computing at Colleges and Universities in the United States Revisited
Lester J. Pourciau, The University of Memphis, USA
Putting Large Volumes of Information on an Intranet
Ann Rockley, The Rockley Group Inc., Canada
Web-supported learning by example
Marco Ronchetti, Dipartimento di Informatica e Studi Aziendali, Italy
Design consideration in the WEQ-Net site development
Katia Barbosa Saikoski & Mara Lucia Fernandes Carneiro, PUCRS - Pontifical Catholic University of Rio Grande do Sul, Brazil
World Wide Web Hypertext Linkage Patterns
Perry Schoon, Illinois State University, USA
ProMediWeb - Medical case training and evaluation using the World Wide Web
Hardy Schulze & Thomas Baehring, University Hospital of Leipzig, Germany; Martin Adler, Sepp Bruckmoser & Martin Fischer, University of Munich, Germany
Risk Assessment and Training about Type-2 Diabetes on the Internet
Hardy Schulze, Thomas Baehring & Stefan R. Bornstein, University Hospital of Leipzig, Germany; Werner A. Scherbaum, University of Düsseldorf, Germany

University Web Management: A Distributed Model
Phyllis C. Self, Scherer Hall, USA
The Probe Method: A Thorough Investigative Approach to Learning
Glenn Shepherd, Eastern Michigan University, USA
Directing Student Web Research: No Surfing Allowed
Karen Smith-Gratto, North Carolina A&T State University, USA; Gloria Edwards, Purdue University, USA
Virtual-U: An Online Learning Environment
Denise Stockley, Chris Groeneboer, Tom Calvert & Linda Harasim, Simon Fraser University, Canada
Cybersearching for a New Career: Exploring Career Hunting in the Electronic Age
Karen Svenningsen, City University of New York, USA
Performance of Bursty World Wide Web (WWW) Sources over ABR
Bobby Vandalore, Shivkumar Kalyanaraman, Raj Jain, Rohit Goyal & Sonia Fahmy, The Ohio State University, USA; Seong-Cheol Kim, Samsung Electronics Co. Ltd., Korea
Develop Your Own Multimedia Projects for the WWW
Shelle VanEtten & Tarrae Bertrand-Hines, University of New Mexico, USA
Electronic Teaching and Learning - New Ways To Learn - Challenge or Menace for Future Education
Jürgen Vaupel & Manfred Sommer, University of Marburg, Germany
Developing an Online Web-based Course
Ronald J. Vetter, University of North Carolina at Wilmington, USA
Internet Technologies Enhance Allied Health Professionals' Knowledge
Kenneth E. Wright, Vivian H. Wright & Michael Newman, The University of Alabama, USA
Setting Up a Web Server For Interactive Engineering Applications
Husam M. Yaghi, Southern University, USA; John M. Tyler, Louisiana State University, USA
The Internet as a Professional Development and Instructional Resource
Bill Yates & Stephanie Salzman, Idaho State University, USA
Constructing Knowledge in Electronics with the Web
Kai-hing Yeung, The Hong Kong Institute of Education, Hong Kong
IP Packet Filtering Interface Design: Providing Fast and Time Predictable Web Infoshop Services
Chang-Woo Yoon, Electronics and Telecommunications Research Institute, Korea
Virginia Commonwealth University Events Calendar
James B. Yucha, Virginia Commonwealth University, USA
Tools for Web Site Management
Greg Zick, Jill Fluvog, Craig Yamashita & Lainey Kahlstrom, University of Washington, USA

SHORT PAPERS

Strategies for Effective Use of the Internet and Email for Instruction
Gregory F. Aloia, Ramesh Chaudhari, Jeffrey P. Bakken & John Beemsterboer, Illinois State University, USA
The Effects of Navigation Maps on World Wide Web Usability
Lara Ashmore, University of Virginia, USA
Creating Web Presence using NetPresence System
Aaron Aw, Chan Fang Khoon, Jimmy Tan, Louisa Chan & Tan Tin Wee, National University of Singapore, Singapore
Development of Simple Campus Intranet Courseware of English as a Foreign Language for Japanese Learners
Junichi Azuma & Kazuhiro Nomura , University of Marketing and Distribution Sciences, Japan
Virtual Workshop: Interactive Web-Based Education
Kathy Barbieri & Susan Mehringer, Cornell University, USA
Psychological Analysis of Specific Development Problem "Users from Former Soviet Union in Internet and WWW Spaces"
Valentina M. Bondarovskaia, Institute of Applied Informatics, Ukraine
The Interactive Learning Connection - University Space Network (ILC-USN). A Hybrid Internet/CD-ROM Course in Spacecraft Systems Design
William Brimley, Ryerson Polytechnic University, Canada

Distance Learning - Courseware Development
Netiva Caftori, Northeastern Illinois University, USA
Making the MOST of Virtual Reality
Margaret D. Corbit, Cornell University, USA
FishNet: Finding and Maintaining Information on the Net
Paul De Bra, Eindhoven University of Technology, The Netherlands & University of Antwerp, Belgium; Pim Lemmens, Eindhoven University of Technology, The Netherlands
Task-centered Navigation in a Web-accessible Data Space
Eckehard Doerry, Arthur E. Kirkpatrick, Sarah Douglas & Monte Westerfield, University of Oregon, USA
Integrating Security Services Into Collaborative Systems
B. S. Doherty & M. A. Maarof, Aston University, United Kingdom
NJECHO - An Electronic Journal focused on Computers Across the Curriculum
Beva Eastman, William Paterson University, USA
Integrating Internet Technology into Distance Teaching at the Open University of Israel
Benny Friedman & Michal Beller, The Open University of Israel, Israel
Facilitating a Multimedia Design Course through Web-based Communication
Jane M. Fritz, University of New Brunswick, Canada
Virtuality of MOOs as a Language Learning Tool
Francesca O. Galante, University of Wisconsin-Madison, USA
Multimedia Pedagogy - Creating longevity in CAL applications!
J.G. Gallagher & D.P. Stevenson, Napier University, Scotland; E. Fordyce, Thames Valley University, England
Brokering the Web
K. Garcha, GPT Limited, United Kingdom
EDU-EX: Design of a Knowledge Based Tool to Create Pedagogical Intelligent Educational Systems in the WEB
Angel García-Crespo, Paloma Domingo, Miryam Lancha, Belén Ruiz Mezcua & Antonio de Amescua, Universidad Carlos III de Madrid, Spain
Directory of Available Projects: A Tool for Students, Facult y, and Administrators
Joanna L. Gaski, Amy L. Marr & Charles Kornik, Worcester Polytechnic Institute, MA, USA
Creating an "Exchange" Through Online Web-based Events
Melanie Goldman, BBN Corporation, USA
Mallard: A Java Enhanced Learning Environment
Charles R. Graham, Michael L. Swafford & Donna J. Brown, University of Illinois at Urbana, USA
Comparison of Two Browser Interfaces: Impact of Browser Migration to On-Going Design
Lee Ann Grim, AT&T, USA
On-line Access to the 17th Century Literature
H. Haddouti & S. Wiesener, Bavarian Research Center for Knowledge-Based Systems (FORWISS), Germany
Construction of Consulting Server
Kenichi Hagiwara, Fuji Electric Co., Ltd., Japan
Student Evaluation of a Web Based Course
Dwayne Harapnuik & T. Craig Montgomerie, University of Alberta, Canada
Questionnaire Administration Via the WWW: A Validation & Reliability Study for a User Satisfaction Questionnaire
Ben D. Harper, Laura A. Slaughter & Kent L. Norman, University of Maryland, USA
The Design, Development, and Implementation of a Virtual Online Classroom: From a Designer's Point of View
Marilyn J. Heath, University of Houston, USA
The Pleasure Principle Revisited in Virtual Reality
Francoise Herrmann, Independent researcher, USA
Herbal-T Internet Desktop
Israel Hilerio, i2 Technologies, USA; Weidong Chen, Southern Methodist University, USA
Using Web Sites in University Courses as Bulletin Boards and for Enrichment Use of Natural Language Technics to Improve Text Retrieval on The Web
C Jacquin & M Liscouet, IUT de Nantes, France

Intranet IC Design Environment
Chau-Iou Jeng, National Semiconductor, USA

BIOME: A Scientific Data Archive Search-and-Order System Using Browser-Aware, Dynamic Pages
Sarah V. Jennings, University of Tennessee, USA; Teresa G. Yow, Oak Ridge National Laboratory, USA; Vickie W. Ng, Hughes Information Technology Systems, USA

Integrating Technology into Secondary Science and Mathematics Classrooms: A Partnership
Elizabeth T. Joseph, Slippery Rock University, USA; Judy Pezzulo, Greenville Area School District, Pennsylvania, USA

HANMAUM - A Multi-agent Model for Customer, Merchant and Directory Service
Dae-Ki Kang, Joong-Bae Kim & Ho-Sang Ham, Systems Engineering Research Institute, South Korea

Integrating Web-based Systems into Teaching Instructional Technology
Edward Kazlauskas & I. Soner Yildirim, University of Southern California, USA; Daniel T. Chan, HYNET Technologies, USA

The Web Lecture System (WLS)
Richard L. Klevans, North Carolina State University, USA

RISK-1: its way to the Virtual Reality (Taking risk in virtual environment)
Vladimir Kosolapov, National Academy of Sciences of Ukraine

The Initial Reaction of Users to CALLware
Eva Fung-kuen Lai, The Chinese University of Hong Kong

Interactive Shared Bookmark
Luigi Lancieri, France Telecom, France

Distributed Administration-Service Integration on the World Wide Web
Carlos Laufer & Hugo Fuks, Catholic University of Rio de Janeiro, Brazil

Teaching College Mathematics with the WEB, Scientific Workplace, and Scientific Notebook
Lawrence E. Levine, Stevens Institute of Technology , USA

The Walls Come Tumbling Down Pedagogical Issues and the Web
James Lockard, Ana Donaldson & Marybeth Koos, Northern Illinois University, USA

Agora: Shaping Social Space from Information Space
Byron Long & Ronald Baecker, University of Toronto, Canada

A Framework for Developing Web-Based Educational Applications: Using a Relational Database to Connect the Design of Authoring and Navigational Tools
Yibiao Lu, Matthew J. Koehler & Richard Lehrer, University of Wisconsin-Madison, USA; Wei-Tek Tsai, University of Minnesota, USA

CoProcess: A Java-based Environment for Collaborative Process Management Over the Web
K. J. Maly, H. Abdel-Wahab, R. Mukkamala, A. Gupta, M. Kholief & S. Dittakavi, Old Dominion University, USA; C.S. Vemuru & H. Syed, Innovative Aerodynamic Technologies, USA

A CAL Package for HTML Basics
Moira J. McAlister, University of Sunderland, United Kingdom

Issues: The Internet and Society
Leah P. McCoy, Wake Forest University, USA

An Authoring Tool for Constructing Interactive Exercises based on Automatic Generation of Client-side Functionality
Sean McKeever, Damien McKeever & John Elder, The Queens Univerity of Belfast, Ireland

Building Context-Sensitivity Through Behavioral Prototyping: The Development of an Internet-Based Mentoring Program for 5th-12th Grade Students
Jay Melican & Eli Blevis, Illinois Institute of Technology, USA

A Design Procedure for Training in Web
Ch. Metaxaki- Kossionides, University of Athens, Greece

Dynamic Educational Applications with Multimedia Databases: A Prototype
Sara Moshman, University of Nebraska-Lincoln, USA

OCS - An Online Conference Server
Michael Müldner & Ivan Tomek, Acadia University, Canada

The 15 Minute Series : Tools for Internet Trainers
Robin Murphy, Network Solutions, Inc., USA

WebView: A Multimedia Database Resource Integration and Search System over Web
Deepak Murthy & Aidong Zhang, State University of New York at Buffalo, USA
NetTest: An Integrated Web-Based Test Tool
Yun Ni, Jianping Zhang & Donald H. Cooley, Utah State University, USA
The Quest for Major Forces Underpinning the Internet Worldwide Landscape
Enrico Nicolo' & Bartolomeo Sapio, Fondazione Ugo Bordoni, Italy
Distributed Curriculum Authoring Using SGML
Roger Nkambou & Gilles Gauthier, Université du Québec à Montréal, Canada
Exploring Culture on the Web
Maritza M. Osuna & Carla Meskill, University at Albany - SUNYA- New York, USA
Web Access to CD-ROM Databases: A Model for Easy, Equitable Access to Government Information
Carolyn Ottow, Oregon State University, USA
The Intranet Support Problem
John R. Potter, John Deere Waterloo Works, USA
Logan: Formal Telemeeting Document Support for the World-Wide Web
Gitesh K. Raikundalia, Bond University, Australia
InterWeb : A Web-Based System for Collaborative Learning and Work
Riccardo Rizzo & Santi Scimeca, Istituto di Tecnologie Didattiche e Formative, Italy
Electronic Transcripts in the Age of Standards
Robert Roberts, Oregon State System of Higher Education, USA; Robby Robson, Oregon State University, USA
Productivity Tools for Creating Web-based Presentations
William Rosener, Northeastern State University, USA; Julie Little-McGill & John Rose, University of Tennessee, USA
Nääsnetti - a 3D Media Village for Social and Informational Uses of Media
Timo Saari, Aamulehti Group Ltd., Finland
A Web-Based Training and Advising System for Software Libraries
Ahmed Seffah, Computer Research Institute of Montreal, Canada; Cao Lieu Nguyen, Université du Québec à Montréal, Canada
Designers Designing a Web Site to Teach Designers to Design for the Web
Cary Staples, University of Tennessee-Knoxville, USA; Jamison Abart & Tim Parker, Visuoso Media, USA
Using Internet Technology and Performance Support Methodology for Teacher Training
Ginger Swope, George Huffman & Phyllis McVoy, Florida State University, USA
Flora of North America: A Distributed Cognitive System
Kay L. Tomlinson, Mark A. Spasser & John L. Schnase, Missouri Botanical Garden, USA
WASE: Towards a Web-Aided Software Engineering
Ju An Wang, University of Macau, Macau
Roadmap to ATM
Glenn Warnock, Douglas Reid & Maike Miller, Algonquin College, Canada
A Distance and Open Learning Project for Multimedia Training
Gérard Weidenfeld, Université de Picardie Jules Verne, France
The Effects of Waiting Time on Website Utilization
Bruce D. Weinberg, Boston University, USA
Education on the Net: The Experience of the Writers in Electronic Residence Program
Herbert H. Wideman, York University, Canada
Overcoming Electronic Course Delivery's Greatest Obstacle: Specific Policy Recommendations for Institutions of Higher Learning
David Wiley, Marshall University, USA

SHORT PAPERS (Works In Progress)

Supporting Collaborative Work in Intranet Using Active DBMS Implemented with Java
Philippe Aniorté, Marc Dalmau & Philippe Roose, IUT Informatique de Bayonne, France
Net-based Open & Distance Learning: Experiences from the MECPOL project (Models for European Collaboration and Pedagogy in Open Learning)
Bodil Ask, Agder College, Norway; Harald Haugen, Stord/Haugesund College, Norway

Team Development Of Web-Based Applications: Experiences In A Software Engineering Course
George E. Atkins, Southwestern Oklahoma State University, USA

Designing Hybrid CD-ROM/Web: A Reality Check
William G. Barek, University of Toronto, Canada

"Hangar Flying" as story-based instruction: Capturing expertise via online video libraries
Benjamin Bell, Sholom Gold & Danielle Kaplan, Columbia University, USA

Bartering and Gaming for Education on the Web with Non-convertable Virtual Currency
Gary McI. Boyd, Concordia University, Canada

An Interactive, Interdisciplinary Web Site Template for Elementary Education Using Remotely Sensed Imagery
Ginger Butcher, Jill Twetten & Houri Tamizifar, SGT, Inc., USA; Josephine To, Science Systems and Applications, Inc., USA; Jon Avila, Mary Washington College, USA; Karen Kaye, USGS Biological Resources Division, USA; William Webster, NASA Goddard Space Flight Center, USA

An Aglet-based Distributed Web Indexer
Marco Chirico, Fabrizio Giudici, Andrea Sappia & Anna Marina Scapolla, University of Genoa, Italy

A Monitor System for Cisco Router Interfaces: A Case Study for a University Metropolitan Area Network
Enrico Commis, Università di Catania, Italy; Luca Gentili, CINECA, Italy

Computational Agent to Integrate An Expert System for Design and the Internet
Jonny Carlos da Silva, Federal University of Santa Catarina, Brazil; David Dawson, Lancaster University, United Kingdom

A Simple Model for Adaptive Courseware Navigation
D. Pilar da Silva, R. Van Durm, K. Hendrikx & E. Duval, H. Olivié, Katholieke Universiteit Leuven, Belgium

Virtual Institute Center for the advancement of two-year to four-year transfer
JR Garcia de Alba, Texas A & M University, USA

An Adaptive Multi-Agent Architecture for the ProFusion* Meta Search System
Yizhong Fan, Sprint Corporation, USA; Susan Gauch, University of Kansas, USA

Using Active Filters to Improve Foreign Language Instruction
Virginia M. Fichera, Doug Lea & Joseph Grieco, State University of New York at Oswego, USA

Instructional Design Issues for Web-based Instruction:Teaching an Old Model New Tricks
Kristina A. Fritts & Cheryl Krawchuk, University of Arkansas, Fayetteville, USA

A Web-Based Virtual Reality Navigation Research Tool
Greg Furness, Caterpillar, Inc., USA; Joaquin Vila & Barbara Beccue, Illinois State University, USA

Dynamic Construction of Geographical Hyperdocuments
Mauro Gaio, Maurice Szmurlo & Jacques Madelaine, University of Caen, France

Targeting User Needs with Customized Web Views
Kathryn F. Gates, University of Mississippi, USA

Roadblocks and Breakthroughs in Integrating the WWW into the University Curriculum
Fred Herschede & E. Marcia Sheridan, Indiana University South Bend, USA

Issues in Distance Education
Ali Jafari, Ph.D., Indiana University, USA

A Form-based Gateway for Web Databases
Euna Jeong & Jane Yung-jen Hsu, National Taiwan University, Taiwan

Web Server Performance and Process Model
Tae-sung Kim & Hyuck Yoo, Korea University, Korea

Infusing Web Content with Educationally Relevant Indices
Rose Kotwas, Lincoln High School, USA; Janet Luongo & Ben Ortiz, The Discovery Museum, USA; Miriam Masullo, Joan Musgrave & Joanne Tursky, IBM T.J. Watson Research Center, USA; Ray Ramirez, North Central Regional Educational Laboratory, USA

Towards a Web Operating System (WOS)
Peter Kropf, Université Laval, Canada; John Plaice, University of New South Wales, Australia; Herwig Unger, Universität Rostock, Germany

An User Authentication Method for WWW using Key Distribution Algorithm and Random Sequence
Yun-ho Lee & Young-sun Kim, Korea Telecom R&D Group, Republic of Korea

Using Web Technologies to Support a Knowledge Management Culture
Paul Lefrere & Andy Stone, Open University, United Kingdom
NDSU's Tailor-made Web site Revamp
J. Ludvik Herrera Lopez, North Dakota State University, USA
Intelligent Indexation of Compound Document using Images
Nicolas Malandain, Mauro Gaio & Maurice Szmurlo, University of Caen, France
The Virtual Moderator™ Project
Jacques Malouin, University of Sherbrooke, Canada
Interactive Computer Ethics Explorer
Walter Maner, Bowling Green State University, USA
A Group Interaction System for Web-Based Market Places
Peter Manhart, Karin Schmidt, Johannes Bumiller & Felix Meyer, Daimler-Benz Research and Technology, Germany
Searching for Intelligence on the Web: Exploring the use of Multiple Intelligence Theory in the Internet Environment
Ana Beatriz Martinez & Kristina Ann Fritts, University of Arkansas, USA
Enhancing Commercial Web Sites by Dynamic User-Interaction-Management
Felix Meyer, Daimler-Benz Research and Technology, Germany
Customizing the Content of Courseware
Uwe Neuhaus & Gunter Schlageter, Distance University of Hagen, Germany
Access to Web-based Special Education
Kevin K. Nguyen, University of Toronto, Canada
Connecting Business and Education Through the World Wide Web
Gian Pagnucci & William Macauley, Indiana University of Pennsylvania, USA
Architecture for Intranet Collaborative Learning Activities
Joaquim Sousa Pinto, Joaquim Arnaldo Martins & Conceição Baptista Tavares, University of Aveiro / INESC, Portugal
Publishing Complex Network Information on the Intranet
Veikko Punkka, Nokia Telecommunications, Finland
Chinese Information Access Through Internet
Lu Qin, Lee Kin Hong & Yao Jian, The Chinese University of Hong Kong, Hong Kong
Virtual Worlds of Today, Virtual Worlds of Tomorrow
Howard Rosenberg, Rosenberg & Associates, USA
A Dynamic Visualization of Dialog Contents in a Collaborative Communication
Hajime Saitoh & Takashi Maeda, Graduate School of Hokkaido Information University, Japan
Designing a Context-Specific Questionnaire for Online Teaching
Denise Stockley & Vivian Rossner, Simon Fraser University, Canada
A web based automated advisor for delivering purchasing advice
Sandor Szego, Northwestern University, USA
Collaborative Tasks and Outcomes in Online Training: The Infoshare Module
Lucio Teles & Xinchun Wang, Simon Fraser University, Canada
Computer-Mediated Communication in Enhancing Communicative Dialogues on the Web
Seppo Tella & Marja Mononen-Aaltonen, University of Helsinki, Finland
Distributed Web Authoring
Harold Thimbleby, Middlesex University, United Kingdom
Web-based Virtual Learning Environments: Experiences and Futures
Morten Wagner, Jørgen Bøegh & Allan Meng Krebs, DELTA Danish Electronics, Denmark
Conceptualizing a Tri-Modal Model for Web-Based Distance Education
Joel Wilkinson & Jonathan L. Ross, The University of Calgary, Canada
jInput - A Java applet for CJK keyboard input method
Leong Kok Yong, Liu Hai & Oliver Wu Peng, National University of Singapore, Singapore
Visualization Tool for Collaborative Web Browsing
Guillermo S. Zeballos & Marian G. Williams, University of Massachusetts Lowell, USA

Tiger: Educational Research in the Age of the Internet
Yong Zhao, Valerie L. Worthington & Richard E. Ferdig, Michigan State University, USA

SIG DISCUSSIONS

Using the Internet for and by Teachers
Susan Johnson, Ankeny High School, USA; Bill Rudolph & Janet Sharp, Iowa State University, USA

Web Based Course Development: Advantages and Concerns
Charles H. McCloy & Nancy Simpson, Trident Technical College, USA

AWARD PAPERS

Implementing A Digital Signature Infrastructure for Web Applications

Lalitha Bhat

Internet Center, Bell Atlantic, USA

lbhat@iscorp.bellatlantic.com

Chencheng Qiu

Internet Center, Bell Atlantic, USA

cqiu@iscorp.bellatlantic.com

Abstract: *This paper describes an implementation of a Digital Signature Infrastructure(DSI) using the Yaksha algorithm[Ganesan and Yacobi, 1994], a variant of RSA algorithm that generates signatures identical to RSA signatures. The focus of this DSI system is to provide an easy way to develop web applications with digital signature capabilities. The DSI also provides a simple mechanism to incorporate digital signature functionality for pre-existing web applications.*

Keywords: *Web, Digital Signature, RSA, Yaksha*

Introduction

The wide-spread use of internet/intranet technologies has imposed new requirements on web applications. One important requirement is to authenticate the application data. Digital authorizations, vouchers, receipts etc need proof for their origins in the same way as their paper-based versions need endorser's signature. Technologies such as smart cards are available for digital signatures, yet they have not been widely adopted due to the overwhelming cost.

To fill the gap, we designed and implemented a Digital Signature Infrastructure (DSI) with the Yaksha algorithm. The DSI back-end servers can generate and split RSA keys, certify and store public keys and complete signatures. We also implemented a platform-independent component that the client side of a web application can use to generate partial signatures. The DSI system's architecture is independent of the application that uses it; so not only can new applications take advantage of the infrastructure, but also the existing web applications can easily add digital signature functionality.

The theory of the DSI is based on a variant of RSA algorithm, called "Yaksha". In contrast to the traditional way of storing users' keys on smart cards or on local disk with password protection, Yaksha splits the RSA private key in two parts: a short, easy-to-remember, user supplied key (password) and a longer, system derived key which is stored in a secure server. Both the short key and long key have to work together to complete a signature. In the event that one piece is compromised, the other piece can be instantly revoked.

In the following sections, we first review the RSA and Yaksha signature algorithms. Then we describe the architecture of the Digital Signature Infrastructure which is followed by the implementation of the DSI servers and client. The final section gives the conclusion.

Review of RSA Signature Algorithm

The RSA Signature algorithm [Rivest, Shamir and Adelman, 1978] is based on public key (asymmetric) cryptography. Each user/entity (denoted by *c*) has a pair of keys:

A private key, *(Dc Nc)*, accessible only to the user/entity,

A public key, *(Ec,Nc)*, publicly known to the rest of the world.

If treated as numbers, the two keys satisfy the following mathematical criteria:

For any positive number M, $(M^{Dc} \, mod \, Nc)^{Ec} \, mod \, Nc = (M^{Ec} \, mod \, Nc)^{Dc} \, mod \, Nc = M$ [Cormen, Leiserson and Rivest, 1990, page 831-836]

To sign a message (or its hashing digest [Schneir, 1994]) M, the signer uses the private key and performs the operation

$S = (M^{Dc}) \, mod \, Nc$

to generate a signature S from M. The signer then sends both the M and S to the receiver.

Let M' and S' denote the message and signature that the receiver obtains at the receiving end. The receiver uses the sender's public key and performs the operation

$(S'^{Ec}) \, mod \, Nc = ((M^{Dc}) \, mod \, Nc)^{Ec} \, mod \, Nc = M$

The message M derived from S' should be mathematically equal to the message received, M', if both the message and signature have not been tampered with. If either the message or the signature was corrupted, or the sender forged the signature with an incorrect private key (Dc, Nc), the verification will fail, i.e., M' will not be equal to M.

Overview of Yaksha Signature Algorithm

The Yaksha digital signature algorithm [Ganesan and Yacobi, 1994; Ganesan, 1996; Ganesan, 1995] is a variant of the RSA signature algorithm. Yaksha still maintains the RSA concept of private and public keys for each user, but goes further and splits the RSA private key, Dc, into 2 parts:

A short user private key *Dcu* which the user uses for signature. This key is not stored anywhere in the system.

A longer "Yaksha" private key, *Dcy* which is derived from *Dcu* and *Dc* in a way that the modular multiplication of *Dcu* and *Dcy* equals *Dc*, i.e., $Dcu \times Dcy = Dc \, mod \, (Phi(Nc))$, where *Phi(Nc)* is the Euler's phi function of *Nc*. [Cormen, Leiserson and Rivest, 1990, P817]. This part of the private key is stored in Yaksha signature server.

The user and the Yaksha signature server collaborate to sign a message. User partially signs the message *M* using user private key *Dcu* and calculating

$Sp = M^{Dcu} \, mod \, Nc$

The signature server locates the user Yaksha key (the other part of the user's private key), *Dcy*, and completes the signature

$$S = Sp^{Dcy} \bmod Nc = (M^{Dcu} \bmod Nc)^{Dcy} \bmod Nc = M^{Dcu*Dcy} \bmod Nc = M^{Dc} \bmod Nc$$

The signature S obtained via this process is identical to an RSA signature, and can be verified using the RSA verification algorithm, as described earlier.

Ganesan and Yacobi [Ganesan and Yacobi, 1994] have mathematically proved that the Yaksha algorithm is as secure as RSA and breaking the Yaksha cryptosystem is equivalent to breaking RSA, even in the presence of active adversary. In addition to its security, Yaksha has the following advantages relevant to the digital signature:

1. Easy to memorize RSA keys are too long to be memorized. As a result, any user who wants to use the RSA signature algorithm needs either a smart card, or a private key stored on a local disk with password protection. With the Yaksha signature algorithm, the user signature key, D_{cu}, can be short and hence can be memorized.
2. Instant revocation capability Due to the hierarchical nature of public key certificate management, revocation takes anywhere from a few hours to a few days, since there is no good method for revocation list distribution. This delay is unacceptable in applications where the ability to immediately revoke a certificate is essential (such as calling cards, credit cards etc., where a delay in revocation can make a card issuer vulnerable to significant financial loss). Yaksha keys can be instantly revoked by contacting the Yaksha Signature server, which can invalidate the server part of the stolen key instantly, thus rendering them useless.
3. Better audit facility Yaksha provides easy management of audit trails, since each signature has to go through the Yaksha signature server for completion.

System Architecture

The Digital Signature Infrastructure consists of four multi-threaded servers and a downloadable Java package. The back-end servers provide services for creating, verifying and managing signatures. These services include key generation and splitting, certificate issuing and key storage and retrieval. But most importantly, they provide the interface to the functionalities of digital signature completion and verification.

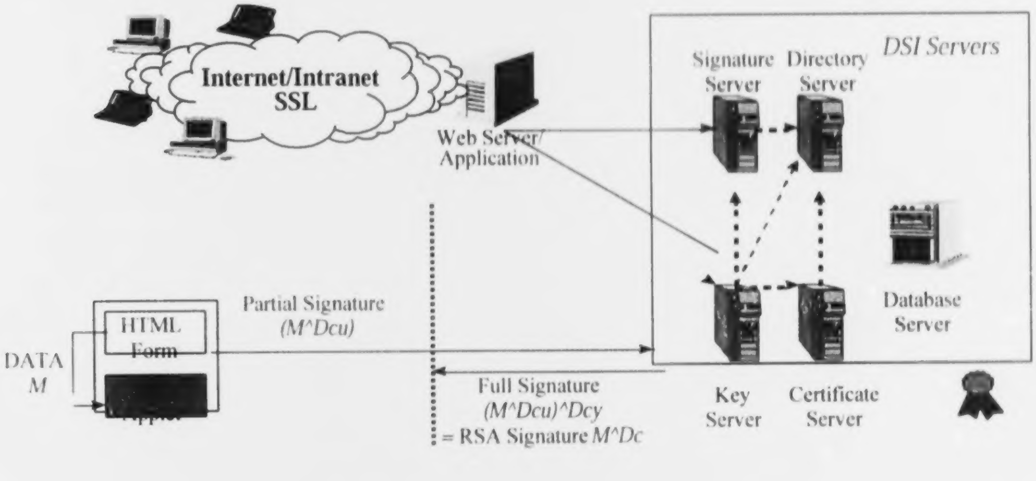


Figure 1. Digital Signature Infrastructure (DSI) Architecture

Back-end servers only store the Yaksha private key *Dcy*, but not user private key or password *Dcu*. To achieve a complete signature, user must partially sign the message using *Dcu*, which is then completed by DSI server using *Dcy*. Because of the insecure nature of the Internet, users' passwords should never be transferred over the Internet, nor should they be stored on the client machine. In our DSI system, they only exist in users' minds. To sign a HTML form data message, a local process (Java applet) asks the user to type in the password, performs the necessary mathematical computations and produces the partial signature. Once the signature is generated, the password can be purged from memory. A later section will describe how the Java applet is implemented to make it easy for web applications to attach digital signatures.

The security of RSA algorithm is based on the fact that it is practically impossible to derive a private key from its corresponding public key. Similarly, in Yaksha, the two pieces of a private key do not convey information about each other. Therefore, DSI servers can not derive user's password Dcu from the Yaksha key Dcy. However, when a partial signature is sent to the DSI signature server for completion, the server can verify user's password by completing the signature and doing a RSA verification. (This is possible because the two-step Yaksha signing produces a RSA signature). If the RSA verification fails, the partial signature is incorrect, possibly caused by an incorrect password provided by the user.

Implementation of Servers

The DSI has four types of servers that interact together to provide all the infrastructure services. The servers are multi-threaded and fully replicable to distribute transactional load. The implementation of servers needs cryptographic operations on large numbers. Commercial off-the-shelf packages such as RSA's Bsafe and Bcert and Bellcore's Large Integer Package (LIP) [Lenstra, 1989] are linked to the system to provide these functions. Following is a brief description of the different servers.

Key Server

The Key Server is responsible for generating RSA keys and splitting them into Yaksha keys. The keys are necessary to establish a person's identity in the DSI, which requires that the user must generate keys as the first step, before signing any document. During the registration process, the user chooses a user ID and password and the key server generates an RSA key pair, (Dc, Ec). It then splits the private key Dc into two parts, the user signature key Dcu (the user-supplied password) and Dcy, the user server key. Since $Dcy = Dcu^{1} * Dc \bmod Phi(Nc)$, the split assumes Dcu^{1}, the modular inverse of *Dcu*, exists for a given *Nc*. If the inverse does not exist for the given RSA key pair and the user-supplied password *Dcu*, new RSA key pairs are generated and the split attempt is repeated.

The strength of the user-picked password is a concern for the security of Yaksha system. An easy to guess password will compromise user's private key, and hence the authenticity of his/her signatures. To mitigate the potential security threat from a poorly picked password, Key Server implements a password filter to prevent "weak" passwords from being used as *Dcu*. The algorithm is based on a dictionary of the probability distribution of characters in possible weak passwords.

After generating three Yaksha key pieces: *Dcu*, *Dcy* and *Ec*, the Key Server interfaces with the Certificate Server to generate the Certificate *Cert* and requests the Directory Server and Signature Server to store the certificate *Cert* and the Yaksha key *Dcy* respectively.

Certificate Server

The Certificate Server receives requests for generating certificates from the Key Server in PKCS #10 format [RSA Laboratories, 1993] and generates an X.509 compatible certificate [CCITT, 1988]. The Certificate Server also serves as a Certificate Authority(CA). It applies the CA's signature using CA's private key to a user public key, thus certifying that the public key belongs to the specific user. When a certificate is generated, it is not stored in the Certificate Server. It is customary to store users' certificates in a directory server using industry standards such as X.500 or LDAP[Yeong, Howes and Kille, 1995]. In our system too, certificates are sent to the Directory Server for storage.

Signature Server

User server keys, D_{cy}, are stored securely in the Signature Server after being generated by the Key Server. All keys reside in a Commercial Database and are encrypted with a symmetric secret key. Only the Signature Server can decrypt them. This ensures that attackers can not get hold of Dcy even if the database is compromised.

Applications send partially signed messages to the Signature Server to be completed. So the primary task of the Signature Server is to complete signatures with user's Dcy, and to verify complete signatures with user's public keys. In fact, when the Signature Server completes a partial signature, it also performs a verification on the completed signature with the user's public key. In this way, it can detect an incorrect password used in partial signature even though it does not know what the password is.

Since all signature requests must go through this server, it is easy to maintain an audit trail of all the signature requests completed or failed. This is one of the benefits of using the Yaksha system. In addition, the Signature Server can instantly suspend a user's Dcy in case of password leakage. Once suspended, the key can no longer be used for signing. This mechanism serves the same purpose as the Certificate Revocation List (CRL) of traditional public key systems, but provides instant revocation capability to better protect users from losses resulting from key compromises.

Directory Server

The Directory Server stores the public key certificates in a database and retrieves them when requested. It is necessary to allow users to change their public key credentials because users may lose passwords, keys may be compromised, or organizational policies enforce regular password change. As a result, the Directory Server may store multiple certificates for a user. While only the most recent certificate is used to sign digital signatures, previously signed documents can be verified by retrieving the certificate that contains the user's public key at the time of the signature.

At the time of our implementation, there were no commercial Directory Servers that complied with current industry standards such as LDAP. So, proprietary format and protocol were used for certificate storage and retrieval. But it is fairly easy to replace it with a commercial Directory Server.

All four DSI servers described above can be running on different machines. This requires that communication among these servers over the network should be authenticated and, if necessary encrypted. Standard RSA key pair authentication and encryption was used for this purpose. Also, to generate keys and certificates, DSI servers need cryptographically strong random seeds, which must be very difficult to guess or predict by potential attackers. The time taken to complete a loop of pseudo-random size was used to generate each byte of the random seed to ensure its unpredictability.

Implementation of DSI Web Client

In Yaksha system, the partial signature has to be generated on the client side, the web browser in this case, to avoid exposing user's password across the insecure network. Many approaches can be used to add the cryptographic functionality to the brower-side of the application: browser plug-in, helper applications, Java applet, JavaScripts and so on. Plug-ins and helper applications were immediately opted out because of their platform and browser dependency. Since the cryptographic operations of digital signature involve complex mathematical calculations on big prime numbers, as well as sophisticated algorithms for hashing, block ciphering etc., we chose to implement a Java package instead of embedding JavaScript in HTML documents.

The idea was further enhanced and a generic Java applet independent of the application domain was designed. The applet takes a data stream as the document input and utilizes user's password to generate a partial signature. Such a reusable component benefits application developers by allowing them to focus on the business logic of the application rather than the details of the digital signature implementation. It also benefits application users, since no installation is needed for the applet; it is automatically downloaded along with the HTML page that embeds it.

Once the partial signature is generated by the Java applet, it can be sent over the network by the application to the DSI server to be fully signed. There is no need to encrypt or authenticate this partial signature, although a secure link between the web browser and server helps prevent attackers from getting hold of the document and partial signature and launching a dictionary attack. Such a link has already been provided by protocols such as Secure Socket Layer (SSL).

Notice in our implementation, the Java applet does not try to communicate with the back-end part of the application or DSI servers. All it does is generate partial signatures for user documents. This design dramatically reduces the development overhead of adding DSI interface to any application. Applications can select their own data flow mechanisms without any special consideration to interface with DSI servers and DSI Java applet. For example, it can be a Java application on both front and back ends via IIOP, or it can be a traditional HTML form and CGI/FastCGI program via HTTP. In the latter case, HTML form data can be tunneled and communicated with Java applet through browser supplied facilities such as the LiveConnect from Netscape and VB Scripts from Microsoft. This feature is especially useful when adding digital signature capability to legacy CGI-based system.

At the time of this implementation, JDK 1.1 had not been finalized. So we could not access the security API that it provides for public key operations. Therefore, the client Java package was implemented for digital signature and its associated cryptographic operations in pure Java. The package not only implemented Yaksha signatures (including the signature completion which is done by DSI servers in our system), but also RSA signature and encryption since the fundamentals of the cryptographic operations are the same. We did not use native C/C++ library with Java since it would cause platform dependency problem as well as client installation need. Performance of cryptographic functions in Java was a concern, since we are talking about complex mathematical operations such as exponentials of two big numbers of 100+ decimal digits here. However, the results of performance tests listed in the table below (the tests were conducted on a Pentium machine with 16MB memory, running Windows 95) indicate that the numbers are presumably acceptable to most web applications and confirm that digital signature can be achieved on web applications with little or no performance sacrifice.

Yaksha Partial Sign	Yaksha Full Sign	RSA Sign	RSA Verify	Yaksha Seal	Yaksha Open
0.39 seconds	4.72 seconds	1.43 seconds	0.14 seconds	4.91 seconds	0.39 seconds

Conclusion

The Digital Signature Infrastructure described in this paper has been used in the development of the Online Password Administration System(OPAS) in Bell Atlantic. OPAS is a web-based electronic logonid/account request system. It is designed to replace the "Login ID Request" paper form with digitally signable web-based form. The system successfully improved the security, auditability and processing turnaround time of the request process. While interfacing with the DSI servers and client for digital signature functions, OPAS has its own business logic, authorization work flow and data requirements. It demonstrates the architecture independence of the Digital Signature Infrastructure and that both new and pre-existing web applications can incorporate digital signature functionality in a cost-effective manner.

Acknowledgment

We would like to thank Ravi Ganesan, the inventor of the Yaksha algorithm, for his critical technical guidance throughout the implementation of the system. Our special thanks also go to our reviewers: Rick Austin, John Bolton and Patrick Widener for their insightful comments on the paper. And finally we owe our special acknowledgment to all the Internet Center staff members of Bell Atlantic for their support of our work.

References

1. [Ganesan and Yacobi, 1994] Ganesan, R. and Y. Yacobi. (Oct. 1994) *A Secure Joint Signature and Key Exchange System.* Bellcore TM24531.
2. [Ganesan, 1996] Ganesan, R. (Mar. 1996) *The Yaksha Security System.* Communications of the ACM.
3. [Ganesan, 1995] Ganesan, R. (Apr. 1995) *Yaksha: Augmenting Kerberos with Public Key Cryptography*. IEEE.
4. [Rivest, Shamir and Adelman, 1978] (Jul. 1978) Rivest, R., A. Shamir and L. Adelman, *On Digital Signatures and Public-Key Cryptography.* Communications of the ACM.
5. [Schneir 1994] Schneir, B. (1994) *Applied Cryptography: Protocols, Algorithms and Source Code in C.* John Wiley and Sons, New York.
6. [RSA Laboratories, 1993]. (Nov. 1993) *Public Key Cryptography Standards # 1: Encryption Standard*.
7. [RSA Laboratories, 1993], (Nov. 1993). *Public Key Cryptography Standards # 10: Certification Request Syntax Standard.*
8. [CCITT, 1988] (1988) *Recommendation X.509: The Directory - Authentication Framework*.
9. [Yeong, Howes and Kille, 1995] Yeong, W., Howes, T., and S. Kille. (Mar. 1995). *Lightweight Directory Access Protocol.* RFC 1777, Performance Systems International, University of Michigan, ISODE Consortium.
10. [Lenstra, 1989] Arjen K. Lenstra. (1989). *Signed Arbitrary Length Long Integer Package*.
11. [Cormen, Leiserson and Rivest, 1990] Cormen, Leiserson, Rivest. (1990) *Introduction to Algorithms.* The MIT Press.

In Search of Web Forms: a Neural Agent for Organizing “Reusable” Notes Documents

Alberto Faro
Istituto di Informatica e Telecomunicazioni, Universita' di Catania, Italy
afaro@k200.cdc.unict.it

Daniela Giordano
Education Department, Concordia University, Montreal, Canada
giord@alcor.concordia.ca

Corrado Santoro
Istituto di Informatica e Telecomunicazioni, Universita' di Catania, Italy
santoro@indy.iit.unict.it

Abstract : The paper proposes a neural agent that performs self-organizing classification to assist in searching and contributing to webs of documents, and in the process of documents reuse. By applying the Kohonen self-organizing feature map (SOFM) algorithm to patterns of influence links among documents it is possible to originate clusters of documents that help infer the aspects that such documents implicitly share. The approach complements search techniques based on semantic indexes. The resulting classification is sensitive to the multiple aspects of a document, which may belong to multiple classes with a varying degree, and allows for treating effectively items that typically have a limited life span, either because they are means to the collaborative production of a more complex item, or because they belong to fast evolving domains. The method has been implemented by Lotus Notes Domino Web server for a case-based application in the domain of information systems design.

1. Introduction

Webs of hypermedia documents need support for interactive exploration, to orient the user and to facilitate effective documents retrieval. Among the solutions that have been recently proposed are perspective walls [MacKinlay et al. 91], interactive dynamic maps [Zizi & Lafon 95], dynamic landscapes [Chalmers et al. 96]. Regardless of which specific front-end visualization technique is adopted, the critical issue for effective use of such webs is finding adequate forms of documents organization to reflect the task domain and support different user typologies. In particular, retrieval for the reuse of documents deserves attention whenever reuse is a process integral to the task, as it is in case-based problem solving and in those tasks that involve collaborative production of documents (e.g., design specifications, building shared models, legal agreements).

Documents can be organized with a varying degree of semantic and structural constraints [Wang & Rada 95], nonetheless there are inherent limitations in retrieval based on semantic indexes. Whether the documents are organized in a conventional database or in a hypertext, searches based on keywords are not robust because of the “vocabulary problem”, i.e., the fact that spontaneous word choice for the same domain by different subjects coincides with less than 20% probability [Furnas et al. 87]. This can be ameliorated by techniques for generating particularly sophisticated thesauri [e.g., Chen et al. 96]; however, the problem that remains open is that indexes rarely support the psychological process of flexible framing of contents [Medin &Ross 89], and of perceiving their multiple facets. As a result, the set of documents retrieved after a search often share only a shallow semantics, in which the context that makes a particular document salient tends to be lost.

The paper proposes a classification technique based on a self-organizing mapping of a web of documents (linked by weighted reference relations) into a set of neurons to highlight classes according to topological properties of the original data space [Kohonen 89]. Reference links take into account influence relations among documents. They are generated by the documents' authors, who acknowledge influence relations by creating citation links to other documents when contributing to the web. Reference links are not typed, to avoid incurring in the indexing problems highlighted above, (as the approach of treating a web as a semantic network would entail), and also because research shows that users resist creating and using typed links [Wang & Rada 95]. The goal is to let emerge from a geography of links a classification that:

- takes into account multiple aspects of a document, so that an item can be considered as belonging to more than one class, with a varying degree;
- allows for treating items in the network that typically have a limited life span, because they are means to the collaborative production of a more complex item or because they belong to fast evolving domains;
- facilitates searching the web and orients the process of contributing an item to the document base.

Following the metaphor of conventional "folders", one might think of a folder as representing, more or less explicitly, the aspects shared by the documents contained in it. The assumption of the paper is that "folders" do not have an a priori ontological status, and it attempts to support the processes underlying folders origination and evolution and the placing of documents in multiple folders. This is helpful especially in two situations : 1) when there is a huge quantity of documents to scan (consultation mode) and 2) when an author, or a team, wants to place its document in context (contributing mode).

Section 2 discusses how a self-organizing classification assists in the consultation, reusing and contributing modes of using the web. Section 3 describes how the Kohonen self-organizing feature map (SOFM) algorithm can be applied to patterns of influence links, to originate documents' clusters that help infer aspects implicitly shared by the documents. Section 4 illustrates an implementation of the method by a Lotus Notes Domino Web server and a neural agent performing Kohonen classification, applied to information systems design.

2. Use and Reuse in Self-Organizing Webs

Our scenario, emphasizing web documents retrieval for reuse, is inspired by the case-based reasoning (CBR) paradigm [Kolodner 93], i.e., an approach to problem solving based on finding the best similar "case" matching the current problem, and then adapting it to solve the problem. The new generated case and the "lessons" it conveys can be contributed to the base of cases, which thus learns the new experience and makes it available for future use.

CBR can be considered an effort in the direction of querying the system in cognitively plausible ways, by resorting to sophisticated indexing schemes and to a carefully chosen vocabulary to ensure a proper level of abstraction. In fact, too abstract indexes may collapse the difference among cases and overgeneralize them, thus providing little heuristic power in finding few best matching cases; on the other hand, highly specific index may fail to capture relevant similarities. Although indexing has been criticized as not being a psychologically plausible model of analog retrieval [Thagard & Holyoak 91], still it proves useful whenever the adopted classification scheme is stable and descriptive enough of the problem and of the domain.

A fixed classification scheme, e.g., indexing, can be adequate for the retrieval of documents based on stable categories such as authors, title or date. When the base of documents is fast evolving, because of contents updates, or because documents are temporary means to produce a deliverable in a cooperative setting, more flexible and evolving classification techniques are needed. The required flexibility aims at tracking a classification process that is fundamentally emergent, and at retaining a discriminatory power for the multiple aspects and issues coexisting in a document, or, with a small leap of abstraction, in a "case". For example, the same piece of information may become irrelevant with respect to a problem, but still it retains some value with respect to an issue unforeseen at the time of the document creation; also, the same piece of information could become obsolete or get incorporated in the web in a more refined form, so that discarding the original source or precedent versions is justified. It is therefore apparent the shortcoming of index based retrieval techniques with respect to capturing the temporal dimension of meaning (topicality, obsolescence, evolution with respect to an issue). Moreover, knowing that a document deals with a very specific topic is not enough, because high semantic precision may not be informative on how the topic is addressed (e.g., the contribution's reason).

A complementary approach to symbolic retrieval is proposed, based on the influence links that trace the document evolution, and whose regularities may be used to discover aspects otherwise concealed. By creating web documents linked by references that do not have an explicit semantics, but that only capture strength of influence, it is possible to originate a space that can be dynamically classified by a self-organizing Kohonen network [Kohonen 1989]. Section 3 discusses in detail how the net's topological organization in classes provides an implicit representation of the aspects shared by the documents classified as belonging to that class.

This can support a search and retrieval mechanism based on two main steps: first an item, or a set of items, is identified based on semantic/lexical criteria (e.g., by full text search or conventional indexes) and then it is proposed with the context (i.e., the class) to which it more strongly belongs. The closer classes are also highlighted, to suggest other relevant items or contexts, following a spreading activation mechanism that makes it likely to find the sought information, suggestion, or item in the surrounding of the retrieved documents. One peculiar advantage afforded by the proposed technique is to provide this neighborhood.

Local links certainly assist in understanding better the meaning and context of the retrieved documents, but the mechanism of local exploration is especially recommended when the organizing principle underlying the current class is not evident *yet*. This is likely to occur when document are not stable, or when the user has some difficulties in framing the search problem. As long as the documents' configuration evolves towards more stable forms, local links become less useful in the search process; however, they still play a role in letting the global forms emerge. The evolution towards clearer forms of organization will be determined by the insertion of new elements that will update the preexisting configuration of links.

It must be noted that meaningful global forms emerge (if they do) only when a huge quantity of interrelated documents are available, thus large scale dynamic classification cannot be the sole responsibility of a human processor. In real life, a small scale approximation of this classification process occurs when a problem is framed and solved by incrementally incorporating the suggestions coming from peer reviewing and expert consultations, each highlighting some particular aspect of the problem. The process validity increases when the number of consultation increases, and when everybody is aware of each other suggestions, as in a meeting or brainstorming session. This is quite rare and quite costly, but fortunately CSCW technologies and models now make it possible to collect contributions in a shared electronic environment, in which the role of the above neural agent is justified, also to support the asynchronous sharing of experiences for reuse.

Another issue is that knowledge bases organized as collection of documents may well contain contradictory elements. When detected, contradictions urge forms of documents' organization in which they are resolved (progress). Contradiction can arise because of items "misplacement", or because the item contains errors or misconceptions. The first problem can be solved by a finer classification of the space of documents, or by "migration" of the element to a more appropriate partition of the documents space. The second one can be solved either by document elimination or by amendment, to inhibit the creation of a new class that would be based on faulty hypothesis. Neural classification can assist in managing the documents' space growth, by allowing obsolete documents elimination only when they belong to classes consolidated in stable ontologies, thus keeping the overall web organization stable in spite of the deleted links.

3. Shaping the Web by Neural Classification

In the following a method is illustrated to infer a similarity degree among documents from the information embedded in the references links and to create clusters of related documents. The method starts by asking the author to link every new document to those dealing with a relevant ontology, by using a quantifier I (Influence weight) defined as follows: $I = 0.5$ if the new document takes into account some marginal aspect of the referenced item, $I = 1$ if the new document inherits several important aspects of the referenced item, and $0.5 < I < 1$ for the intermediate situation. Fig.1 shows influence weights placed on the references links.

When a new document is inserted in the web with its reference links, the set of the already existing classes it belongs to is computed. There are several methods to aggregate an element in an existing class, e.g., based on the information exchanged between the new element and the existing ones [Alexander 64]. Here it is adopted a neural approach based on Kohonen self-organizing networks (or maps), which aggregates the documents in classes, not known *a priori*, that preserve a meaningful topological distribution, i.e., the more aspects are shared, the closer the classes [Kohonen 89].

To this aim, we extract from the web the reference graph consisting of documents interconnected by the above influence weights [Fig. 1a]. This is the input space [Fig. 1b] of the Kohonen network, whose number of output neurons must be equal to the number of classes in which we want to classify the documents in the web. The SOFM algorithm operates a classification in which the spatial distance among neurons that represent the classes mirrors the one of the input space [Fig. 1c].

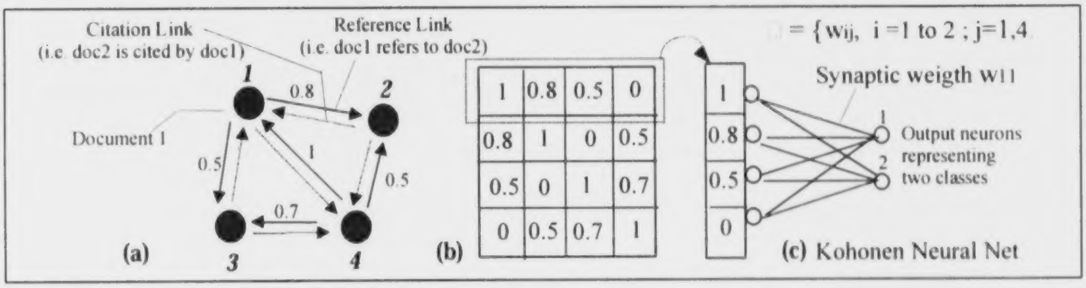

Citation Link
(i.e. doc2 is cited by doc1)
Reference Link
(i.e. doc1 refers to doc2)
Document 1
1
2
3
4
0.8
0.5
1
0.5
0.7
(a)
(b)
1 0.8 0.5 0
0.8 1 0 0.5
0.5 0 1 0.7
0 0.5 0.7 1
□ = {wij, i =1 to 2 ; j=1,4
Synaptic weigth w11
Output neurons representing two classes
(c) Kohonen Neural Net

Figure 1: a) Space of references and citations in the web; b) influence matrix between documents; and c) self-organizing feature map □ to classify documents in classes (input space = influence matrix, output space = neurons representing the classes).

Fig.2 shows how the method proceeds assuming a binary decomposition scheme. Starting from the initial class containing all the cases (Class1), the SOFM algorithm subdivides it in two classes and then re-subdivides Class1.1 and Class1.2 in other two classes and so on. Thus, to classify a new document it is sufficient to start from the class that contains all the items referenced by the new one. For example, if the new document refers to items in Class1.2.1 and Class1.2.2 the method restarts classification from Class1.2.

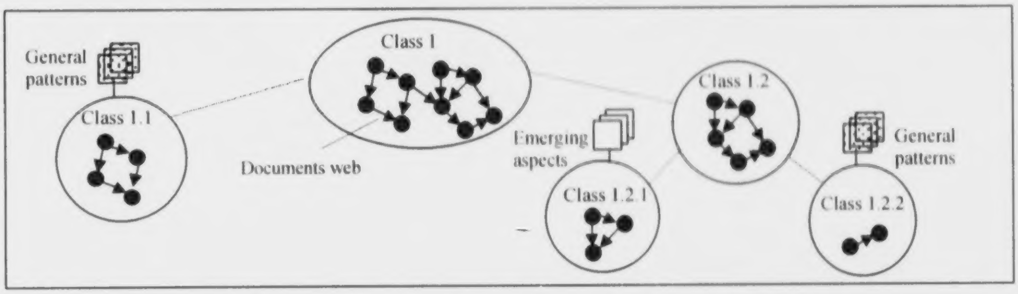

General patterns
Class 1
Class 1.1
Documents web
Class 1.2
Emerging aspects
Class 1.2.1
General patterns
Class 1.2.2

Figure 2: Classification of the documents in the web by the SOFM algorithm.

Experimental evaluation of the SOFM algorithm that we have implemented has shown that binary decomposition of the initial class into 2^k classes (after k successive refinements) is far more accurate than the one step classification obtained by using a Kohonen network with 2^k output neurons. Depth of classification, e.g., the number of levels, can be fixed by the user. In any case, classification is stopped when all the subclasses cannot be further subdivided due to their high interconnection degree (lowest level classes).

The neural classification is repeated every time a new document enters the web; the classes are created and dynamically refined with the web evolution. If the new document does not belong to any existing class, the author is invited to introduce a general description (pattern) to provide some clues concerning the meaning of the newly created class. If the document is placed on an existing class, but the author does not agree with the proposed patterns, s/he can add a new version of the patterns that presently denote the class. If the document belongs to a class not denoted by a pattern yet, the author is "challenged" to identify a general pattern, which is likely to emerge if all the documents referenced by the new one with $I > 0.8$ belong to the same class. Other outputs of the classification are: for each class, a measure of the interconnectedness of the elements in the class (aggregation factor) and, for each element, the degree with which it belongs to all the existing classes.

Adding a new document could modify the structure of the existing classes, i.e., some old document could pass from a class to a different one. However, this phenomenon involves only few documents of the existing classes, and modifies only marginally the structure of the classes. This happens because as long as classes become consolidated, the links introduced by the new item are significantly less in number with respect to the existing ones. The documents that migrate to new or different classes are important to give rise to new ontologies or to reinforce the existing ones. At the end of the decomposition, we have these types of classes:

classes that are denoted by ontological descriptions that give form to the web (e.g., class 1.1or class 1.2.2

in fig.2); such classes are characterized by a high aggregation factor.

classes that cannot be denoted by a single description, either because there is no underlying ontology or because their ontology is so ill-structured that it cannot be expressed explicitly (e.g., class 1.2 in fig.2).

classes that are denoted by partial descriptions pointing out particular aspects that can be taken into account when authoring documents that will be aggregated in the same or the neighboring classes (e.g., class 1.2.1 in fig.2); such classes are characterized by an intermediate aggregation factor.

4. Self-organizing Documents Webs in a Lotus Notes Based Environment

A Lotus Notes based environment, called StoryNet, for the collaborative production of documents structured in stories and episodes, has been enhanced by a neural agent performing classification according to the SOFM algorithm previously outlined. StoryNet's architecture has been conceived to manage evolving systems, and is proving useful for IS collaborative design. The rationale for story based organization is that in such a format experiences can be represented and recollected [Bruner 90]. In the application of StoryNet for IS design, a project consists of a set of use stories and episodes. Each episode is linked to the ones it refers to, and may be reused for specifying analogous episodes. The episode's categories (title, assumptions, what, who, why, when, where, rituals, how, what can go wrong, exception handling) are used as a probe to extract the episodes that best fit the specific design needs [Faro & Giordano 96]. After adapting these episodes, the designer inserts the new episodes in StoryNet. To support reuse, any new document should be inserted as a motivated evolution of the previous ones, i.e., as an enhancement of the experience already captured in the knowledge base. This can be pointed out in comments mediating the references links.

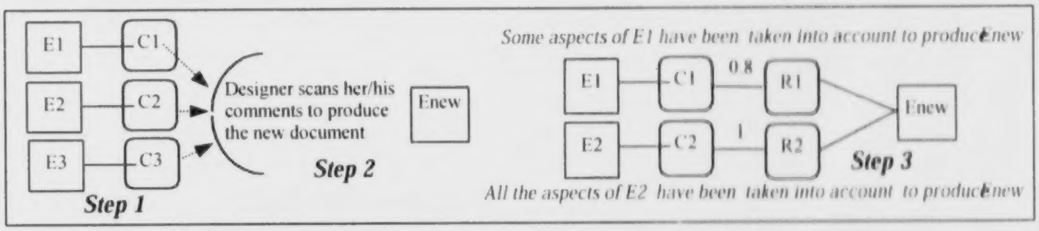
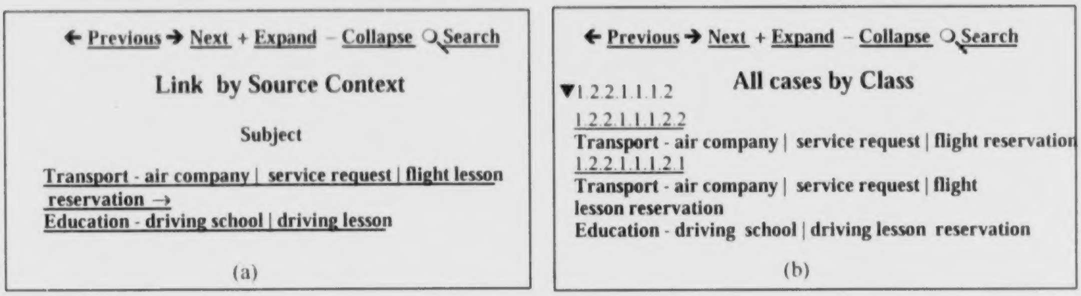

E1
C1
E2
C2
E3
C3
Step 1
Designer scans her/his comments to produce the new document
Enew
Step 2
Some aspects of E1 have been taken into account to produce Enew
E1
C1
0.8
R1
E2
C2
1
R2
Enew
Step 3
All the aspects of E2 have been taken into account to produce Enew

Figure 3: Supporting the referencing process in Storynet (Ei = episode, Ci = comment, Ri = reference).

StoryNet has been implemented by Lotus Notes Domino Web server, to afford easy access to the designer, without requiring a Lotus Notes client. The story-episode organization is easily supported by the Domino Web server, as it is full text search on all the documents. To link episodes belonging to different stories it is necessary to extend the Domino Web server by a suitable software library of C modules that supports the referencing process as follows : 1) the designer first creates special documents to comment the episodes that have been proposed as the result of the search; tipically only the subset considered potentially relevant to the current purposes is marked by a comment [fig.3, step 1]; 2) while detailing the new episodes, the designer may scan the comments for possible suggestions [fig.3, step 2]; 3) after having specified the episode, the designer creates references to the comments that were taken into account [fig.3, step 3]. Passing from an episode to its comments and to its references is supported by the Domino Web server facilities; passing from an episode to the referred ones is supported by the above extension. For example, to pass from a Enew to its referenced items one can obtain the list of all the references, i.e., R1and R2 [fig.3], then pass from0 Ri to Ci by simply clicking a special field inside the reference Ri. After reaching Ci it is easy to pass to episode Ei by the Lotus Notes facilities. Fig. 4 shows how the user can navigate from a document, e.g., "driving lesson reservation", to its source, e.g., "flight lesson reservation", via a reference link. Episodes are organized in a graph whose oriented arcs are labelled by a number measuring how much an existing episode has influenced the new one. The graph is put in a inter-episodes influence matrix stored into a file external to StoryNet, to be elaborated at regular intervals by the neural agent. The agent stores the hierarchical classification of the episode into another file, so that StoryNet can superimpose this classification scheme on the existing episodes. The current version of StoryNet labels each episode by the lowest level class it mainly belongs to, and provides all the classes the episodes belong to.


← Previous → Next + Expand – Collapse Search
(a)
← Previous → Next + Expand – Collapse Search
(b)

Link by Source Context

Subject

Transport - air company | service request | flight lesson reservation →
Education - driving school | driving lesson

All cases by Class

▼1.2.2.1.1.1.2
1.2.2.1.1.1.2.2
Transport - air company | service request | flight reservatio
1.2.2.1.1.1.2.1
Transport - air company | service request | flight lesson reservation
Education - driving school | driving lesson reservation

Figure 4 : (a) StoryNet reference links; (b) StoryNet classification performed by the neural agent

Figure 4b shows the StoryNet user interface for the classification results. Note that "driving lesson reservation" and "flight lesson reservation" belong to the same class, due to the reference link. If an episodes belongs to a class with a degree greater than 0.8, the two relevánt lowest level classes are shown too.

5. Current and Future work

The neural agent implemented in StoryNet has been tested on a small scale set of documents produced for information systems specification, and has generated classifications deemed plausible and useful for guiding searches. We are currently working at large scale testing in the setting of collaborative design assisted by webs of design cases; at testing heuristics for deploying the links and policies for document elimination. Of concern is to what extent the effort of placing the links is offset by the expected gain (individual and collective) in searching and building evolving organizational memories. Another comprehensive test will be performed on large sets of research documents where citation links are already available, such as in the Social Science Citation Index. If performance of the neural agent will be satisfactory, the next step is to find more effective visualization techniques to reflect class topologies and for highlighting items belonging to multiple classes.

6. References

[Alexander 64]. Alexander, C. (1964). *Notes on the synthesis of form*. Harvard University Press.

[Bruner 90]. Bruner, J. (1990). *Acts of meaning*. Harvard University Press.

[Chalmers et al. 96]. Chalmers, M., Ingram R., & Pfranger, C. (1996). Adding imageability features to Information displays. *UIST'96, Seattle, Nov. 6-8, 1996*, Association for Computing Machinery, 33-39.

[Chen et al. 96]. Chen, H. et al. (1996). A Parallel Computing Approach to Creating Engineering Concept Spaces for Semantic Retrieval. *IEEE Trans. on Pattern Analysis and machine intelligence*, 18 (8), 771-782.

[Faro & Giordano 96]. Faro, A., & Giordano, D. (1996). Story telling reasoning to learn Information Systems design. *EuroAIED96, European Conference on Artificial Intelligence in Education*, Lisbon.

[Furnas et al. 87]. Furnas, G. et al. (1987). The vocabulary problem in human-system communication. *Communications of the ACM*, 30 (11), 964-971.

[Kolodner 93]. Kolodner, J. (1993). *Case Based Reasoning*. San Mateo, CA: Morgan Kaufmann.

[Kohonen 89]. Kohonen, T. (1989). *Self organization and associative memory*. Springer Verlag.

[MacKinlay et al. 91]. MacKinlay, J.D., Robertson, G.G., & Card, S.K. (1991). The perspective wall : detail and context smoothly integrated. *CHI'91*, Association for Computing Machinery, 173-180.

[Medin & Ross 89]. Medin, D.L., & Ross, B.H. (1989). *The specific character of abstract thought: categorization, problem solving and induction*. In: Advances in the psychology of human intelligence, Vol. 5. Erlbaum, 189-223.

[Thagard & Holyoak 91]. Thagard, P. & Holyoak, K. (1991). Why indexing is the wrong way of thinking about analog retrieval. DARPA Proceedings on Case-Based Reasoning, 1991.

[Wang & Rada 95]. Wang, W, & Rada, R. (1995) Experiences with semantic net based hypermedia. *International Journal of Human-Computer Studies*, 43, 419-439.

[Zizi & Lafon 95]. Zizi M, Beaudouin-Lafon M (1995). Hypermedia exploration with interactive dynamic maps. *International. Journal of Human-Computer Studies*, 43, 441-464.

Beyond Java: An Infrastructure for High-Performance Mobile Code on the World Wide Web

Michael Franz
Department of Information and Computer Science
University of California
Irvine, CA 92697-3425
franz@uci.edu

Abstract: We are building an infrastructure for the platform-independent distribution and execution of high-performance mobile code as a future Internet technology to complement and perhaps eventually succeed Java. Key to our architecture is a representation for mobile code that is based on adaptive compression of syntax trees. Not only is this representation more than twice as dense as Java byte-codes, but it also encodes semantic information on a much higher level. Unlike linear abstract-machine representations such as p-code and Java byte-codes, our format preserves structural information that is directly beneficial for advanced code optimizations.

Our architecture provides fast on-the-fly native-code generation at load time. To increase performance further, a low-priority compilation thread continually re-optimizes the already executing software base in the background. Since this is strictly a re-compilation of already existing code, and since it occurs completely in the background, speed is not cr itical, so that aggressive, albeit slow, optimization techniques can be employed. Upon completion, the previously executing version of the code is supplanted on-the-fly and re-optimization starts over.

Our technology is being made available under the name "Juice", in the form of plug-in extensions for the Netscape Navigator and Microsoft Internet Explorer families of WWW browsers. Each plug-in contains an on-the-fly code-generator that translates Juice-applets into the native code of the target machine. As far as end-users are concerned, there is no discernible difference between Java-applets and Juice-applets, once that the plug-in has been installed, although the underlying technology is very different. The two kinds of applets can coexist on the same WWW page, and even interact with each other through the browser's API. Our work not only demonstrates that executable content need not necessarily be tied to Java technology, but also suggests how Java can be complemented by alternative solutions, and potentially be displaced by something better.

1. Introduction

One of the most beneficial aspects of the rapid expansion of the Internet is that it is driving the deployment of "open" software standards. We are currently witnessing the introduction of a first suite of interoperability standards that is already having far-reaching influences on software architecture, as it simultaneously also marks the transition to a *component model* of software. The new standards, such as *CORBA* (Object Management Group), *COM/OLE* (Microsoft), and *SOM/OpenDoc* (Apple Computer, IBM, Novell), enable software components to interoperate seamlessly, even when they run on different hardware platforms and have been implemented by different manufacturers. Over time, the monolithic application programs of the past will be supplanted by societies of interoperating, but autonomous, components.

It is only logical that the next development step will lead to even further "open-ness", not only freeing components from all dependence upon particular hardware architectures, but also giving them the autonomy to migrate among machines. Instead of executing complex transactions with a distant server by "remote control" over slow communication links, software systems will then be able to send self-contained mobile agents to a server that complete the transactions autonomously on the user's behalf. The inclusion of *executable content* into electronic documents on the World Wide Web already gives us a preview of how powerful the concept of mobile code is, despite the fact that so far only a unidirectional flow of mobile programs from server to client is supported. Distributed systems that are based on freely-moving agents will be even more powerful.

In order to transfer a mobile program between computers based on different processor architectures, some translation of its representation has to occur at some point, unless the mobile program exists in multiple execution formats simultaneously. Although the latter approach seems feasible in the current context of software distribution via CD-ROM, its limits will soon become apparent when low-bandwidth wireless connectivity becomes pervasive. Hence, a compact *universal* representation for mobile code is required. The search for such a universal representation is the subject of much current research [Engler 1996, Inferno, Lindholm et al. 1996], including recent work of the author [Franz & Kistler 1996, Kistler & Franz 1997].

Although Sun Microsystems' *Java* technology is now the de-facto standard for portable "applets" distributed across the Internet, it remains surprisingly simple to provide alternatives to this platform, even within the context of commercial browser software. We have created such an alternative to the Java platform and named it *Juice*. Juice is an extension of the author's earlier research on portable code and on-the-fly code generation[1] [Franz & Ludwig 1991, Franz 1994a, Franz 1994b]. Our current work is significant on two accounts: First, Juice's portability scheme is technologically more advanced than Java's and may lead the way to future mobile-code architectures. Second, the mere existence of Juice demonstrates that Java can be complemented by alternative technologies (and potentially be gradually displaced by something better) with far less effort than most people seem to assume. In fact, once that Juice has been installed on a machine, end-users need not be concerned at all whether the portable software they are using is based on Juice or on Java. In light of this, we question whether the current level of investment in Java technology is justified, in as far as it is based on the assumption that Java has no alternatives.

In the following, we swiftly introduce the mobile code format upon which all of our work is based. We then give an overview of our run-time architecture, which not only provides on-the-fly code generation, but also dynamic code re-optimization in the background. Finally, we report on the current state of our implementation, specifically the availability of an integrated authoring and execution environment for Juice components, and of a family of plug-in extensions for two popular commercial WWW browsers that enable these browsers to execute Juice-based content.

2. An Effective Representation for Mobile Code

Our mobile-code architecture is based on a software distribution format called *slim binaries* [Franz & Kistler 1996] that constitutes a radical departure from traditional software-portability solutions. Unlike the common approach of representing mobile programs as instruction sequences for a virtual machine, an approach taken both with p-code [Nori et al. 1976] as well as with Java byte-code [Lindholm et al. 1996], the slim binary format is instead based on *adaptive compression of syntax trees* [Franz 1994a]. When compiling a source program into a slim binary, it is first translated into a tree-shaped intermediate data structure in memory that abstractly describes the semantic actions of the program (e.g., "add result of left sub-tree to result of right sub-tree"). This data structure is then compressed by identifying and merging isomorphic sub-trees, turning the tree into a directed acyclic graph with shared sub-trees (for example, all occurrences of "x + y" in the program could be mapped onto a single sub-tree that represents the sum of "x" and "y"). The linearized form of this graph constitutes the slim binary format.

In the actual implementation, tree compression and linearization are performed concurrently, using a variant of the classic LZW data-compression algorithm [Welch 1984]. Unlike the general-purpose compression technique described by Welch, however, our algorithm is able to exploit domain knowledge about the internal structure of the syntax tree being compressed. Consequently, it is able to achieve much higher information densities [Fig. 1]. We know of no conventional data-compression algorithm, regardless of whether applied to source code or to object code (for any architecture, including the Java virtual machine), that can yield a program representation as dense as the slim binary format.

---

[1] note that this earlier work on mobile code predates Java by several years

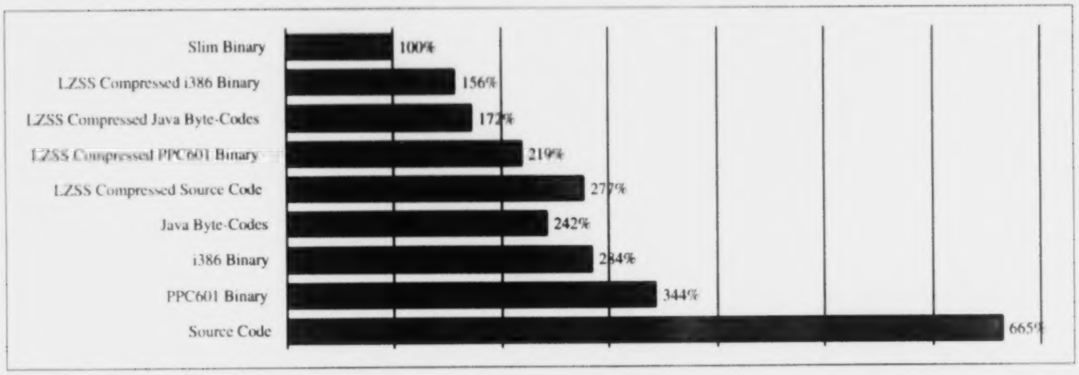


Figure 1: Relative Size of a Representative Program Suite in Various Formats

The compactness of the slim binary format may soon become a major advantage, as many network connections in the near future will be wireless and consequently be restricted to small bandwidths. In such wireless networks, raw throughput rather than network latency again becomes the main bottleneck. We also note that one could abandon native object code altogether in favor of a machine-independent code format if the portable code would not only *run* as fast as native code, but also *start up* just as quickly (implying that there would be no discernible delay for native-code translation). As the author has shown in previous work, this becomes possible if the portable software distribution format is so dense that the additional computational effort required for just-in-time code generation can be compensated entirely by reduced I/O overhead due to much smaller "object files" [Franz 1994a, Franz 1994b, Franz 1997a].

Compactness does come at a small price: since isomorphic sub-trees have been merged during encoding, a program represented in the slim binary format cannot simply be interpreted byte-by-byte. Conversely, the individual symbols in an abstract-machine representation such as Java byte-codes are self-contained, permitting random access to the instruction stream as required for interpreted execution. However, in exchange for giving up the possibility of interpretation, which by its inherent lack of run-time performance is limited to low-end applications anyway, the slim binary format confers a further important advantage:

It turns out that the tree-shaped program representation from which the slim binary format is generated (and which is re-created in memory when a slim binary file is decoded) is an almost perfect input for an optimizing code generator. The slim binary format preserves structural information such as control flow and variable scope that is lost in the transition to linear representations such as Java byte-codes. In order to perform code generation with advanced optimizations from a byte-code representation, a time-consuming pre-processing step is needed to re-create the lost structural information. This is not necessary with slim binaries. A similar argument applies with respect to *code verification*: analyzing a mobile program for violation of type and scoping rules is much simpler when the program has a tree-based representation than it is with a linear byte-code sequence.

3. A Run-Time Architecture Featuring Dynamic Re-Optimization

We are developing a run-time architecture in which the capability of generating executable code from a portable intermediate representation is a *central function of the operating system itself* [Franz 1997b]. It thereby becomes possible to perform advanced optimizations that transcend the boundaries between individual portable components, as well as the boundary between user-level and system-level code.

Consider a scenario in which a user downloads several portable components from various Internet sites during a single computing session. Every time that such a component is downloaded, it is translated on-the-fly into the native code of the target machine so that it will execute efficiently. This "just-in-time" translation is able to achieve remarkable speed-up factors when compared to interpreted execution, but it still cannot extract the theoretically achievable optimum performance from the system as a whole. This is because every component has been compiled and optimized individually, rather than in the context of all other components in the system.

In order to achieve even better performance, one would have to perform inter-component optimizations. Examples of such optimizations are procedure inlining across component boundaries, inter-procedural register allocation, and global cache coordination. However, since the set of participating components is open-ended and the user has the option of interactively adding further components at any time, it is of course impossible to perform these optimizations statically. Unfortunately, the principle of dynamic composability that fundamentally underlies open, component-based systems runs counter to the needs of optimizing compilers. The problem is compounded further by the fact that component-based systems are often made out of a relatively large numbers of relatively small parts.

There is, however, a solution: at any given time, the set of *currently* active components is well known. Hence, a globally optimized version of the system can in fact be constructed, except that this has to be done at run-time and that its validity extends only until the user adds the next component. This leads to the key idea of our run-time architecture: to perform the translation from the slim binary distribution format into executable code not just once, but to do so continually, constructing a *series* of globally cross-optimized code images in memory, each of which encompasses all of the currently loaded components. Whenever such a cross-optimized image has been constructed, it supersedes the previously executing version of the same code, i.e. the new code image is "hot-swapped" into the operational state while the previous one is discarded. At the same time, construction of yet another code image is initiated. We call this iterative process *re-optimization*, and it is performed with low priority in the background.

Since re-optimization occurs in the background while an alternate version of the same software is already executing in the foreground, it is largely irrelevant how long this process takes. This means that far more aggressive optimization strategies can be employed than would be possible in an interactive context. Further, because re-optimization occurs at run-time, "live" execution-profile data can be taken into account for certain optimizations [Ingalls 1971, Hansen 1974, Chang et al. 1991]. This is why our model is continuous: although re-optimization would strictly be necessary only whenever new components are added to the system, usage patterns among the existing components still shift over time. Re-optimization at regular intervals makes it possible to take these shifts into account as well. Our system bases each new code image on dynamic profiling data collected just moments earlier, and hence can provide a level of fine-tuning that is not possible with statically-compiled code.

This leaves the question of what happens when a new component is added interactively to the running system. Clearly, one cannot wait for the completion of a full re-optimization cycle of the whole system before the new component can be used. This problem is taken care of by a second operational mode of our code generator: besides being able to generate high-quality optimized code in the background, it also has a "burst" mode in which compilation speed is put ahead of code quality so that execution can commence immediately. Using this "burst" mode, each new component is translated into native code as a stand-alone piece of code not cross-optimized with the rest of the system. For a short while, it will then execute at less than optimum performance. Upon the next re-optimization cycle, it will automatically be integrated with the remaining system and henceforth run more efficiently.

4. Our Prototype Implementation

Our work has originated and continues to evolve in the context of the *Oberon System* [Wirth & Gutknecht 1989, Wirth & Gutknecht 1992]. Oberon constitutes a highly dynamic software environment in which executing code can be extended by further functionality at run-time. The unit of extensibility in Oberon is the *module*; modules are composed, compiled and distributed separately of each other. Oberon is programmed in a language of the same name [Wirth 1988], a direct successor of Pascal and Modula-2. The Oberon System is available on a wide variety of platforms [Franz 1993, Brandis et al. 1995].

For all practical purposes, Oberon's modules supply exactly the functionality that is required for modeling mobile components. Modules provide encapsulation, their interfaces are type-checked at compilation time and again during linking, and they are an esthetically pleasing language construct. The only feature that we have recently added to the original language definition is a scheme for the globally unique naming of qualified identifiers. Hence, when we have been talking about "components" above, we were referring to Oberon modules.

We have already come quite far in deploying the ideas described above in a broader sense than merely implementing them in a research prototype. The current Oberon software distribution [Oberon] uses the architecture-neutral slim binary format to represent object code across a variety of processors. Our on-the-fly code generators have turned out to be so reliable that the provision of native binaries could be discontinued altogether, resulting in a significantly reduced maintenance overhead for the distribution package. Currently, our implementations for Apple Macintosh on both the *MC680x0* and the *PowerPC* platforms (native on each) and for the *i80x86* platform under Microsoft Windows 95 all share the identical object modules, except for a small machine-specific core that incorporates the respective dynamic code generators and a minimal amount of "glue" to interface with the respective host operating systems.

The latest release of the Oberon software distribution additionally contains an authoring kit for our Juice mobile-component architecture. The main difference between ordinary Oberon modules and Juice components is that they are based on different sets of libraries. The Juice API is smaller than Oberon's, and modeled after Netscape's Java-Applet-API. Components that are based on this reduced system interface cannot only be executed within the Oberon environment, but also within the *Netscape Navigator* and *Microsoft Internet Explorer* families of WWW browsers, both on the Macintosh (PowerPC) and Microsoft Windows (i 80x86) platforms. Hence, by choosing the optional Juice API rather than Oberon's standard libraries, developers of Oberon-based components can address a much larger potential market.

In order to enable Juice components to execute within the aforementioned WWW browsers, we supply a set of platform-specific plug-ins [Juice]. Each plug-in contains a dynamic code-generator that translates the slim binary representation into the native code of the respective target architecture (PowerPC or Intel 80x86). This translation occurs before the applet is started, using the aforementioned "burst mode" of code generation. It is fast enough not to be noticed under normal circumstances, and the resulting code quality is comparable to the current generation of just-in-time Java compilers. Unlike our Oberon-based research platform, our Juice plug-ins do not yet provide background re-optimization and the additional performance gains that come with it. However, we plan to periodically incorporate our research results into Juice.

Juice differs considerably from Java, yet from the web-browsing end-user's perspective, there is no obvious difference between Java and Juice applets. We claim that this is important, because it shows that Java can be complemented by alternative technologies in a user-transparent manner. In the long run, the choice of a particular mobile-code solution may often simply be a matter of personal taste, rather than a technological necessity. Luckily, it is the applet developer that needs to make this choice; the end user need not know any of it as multiple mobile-code technologies, such as Java and Juice, can happily coexist, even on the same web page.

5. Conclusion and Outlook

Mobile code for the Internet need not necessarily be tied to Java technology. In this paper, we have presented various aspects of a mobile-code infrastructure that differs from Java on several key accounts. Not only is our implementation a test-bed for novel code-representation and dynamic-compilation techniques, but it also confirms the suitability of the existing browser plug-in mechanism for supporting alternative software portability solutions.

As our implementation demonstrates, the plug-in mechanism can even be utilized to provide on-the-fly native-code generation, enabling alternative portability schemes to compete head-on with Java in terms of execution speed. Using plug-in extensions for the most popular browsers, many mobile-code formats could potentially be introduced side-by-side over time, *gradually* reducing Java's pre-eminence rather than having to displace it abruptly. This would make the eventual migration path from Java to a successor standard at the end of Java's life-cycle much less painful than most people anticipate now. The same strategy could also be employed to simultaneously support several mutually incompatible enhancements of the original Java standard.

We contend that dynamic code generation technology is reaching a level of maturity that it will soon be relatively inexpensive to support multiple software distribution formats concurrently. It will then become less important how much "market share" any incumbent software distribution format such as Java byte-codes or Intel binary code already owns. In order to be commercially successful, future software distribution formats will have to mimic Java as far as providing architecture neutrality and safety, but further considerations such as code density will surely gain in importance. Some future formats, for instance, will be more narrowly targeted towards particular application domains. In this larger context, the current enthusiasm surrounding Java may soon appear to have been somewhat overblown.

Acknowledgement

The author gratefully acknowledges Thomas Kistler as the co-author of the Juice system. Part of this work is being funded by the National Science Foundation under grant CCR-9701400.

References

[Brandis et al. 1995] M. Brandis, R. Crelier, M. Franz, and J. Templ (1995); "The Oberon System Family"; *Software-Practice and Experience*, 25:12, 1331-1366.

[Chang et al. 1991] P. P. Chang, S. A. Mahlke, and W. W. Hwu (1991); "Using Profile Information to Assist Classic Code Optimizations"; *Software–Practice and Experience*, 21:12, 1301-1321.

[Engler 1996] D. R. Engler (1996); "Vcode: A Retargetable, Extensible, Very Fast Dynamic Code Generation System"; *Proceedings of the ACM Sigplan '96 Conference on Programming Language Design and Implementation* , published as *ACM Sigplan Notices*, 31:5, 160-170.

[Franz 1993] M. Franz (1993); "Emulating an Operating System on Top of Another"; *Software-Practice and Experience*, 23:6, 677-692.

[Franz 1994a] M. Franz (1994); *Code-Generation On-the-Fly: A Key to Portable Software* ; Doctoral Dissertation No. 10497, ETH Zürich, simultaneously published by Verlag der Fachvereine, Zürich, ISBN 3-7281-2115-0.

[Franz 1994b] M. Franz (1994); "Technological Steps toward a Software Component Industry"; in J. Gutknecht (Ed.), *Programming Languages and System Architectures* , Springer Lecture Notes in Computer Science, No. 782, 259-281.

[Franz 1997a] M. Franz (1997); "Dynamic Linking of Software Components"; *IEEE Computer*, 30:3, 74-81.

[Franz 1997b] M. Franz (1997); "Run-Time Code Generation as a Central System Service"; in *The Sixth Workshop on Hot Topics in Operating Systems (HotOS-VI)*, IEEE Computer Society Press, ISBN 0-8186-7834-8, 112-117.

[Franz & Kistler 1996] M. Franz and T. Kistler (1996); "Slim Binaries"; *Communications of the ACM*, to appear; also available as Technical Report No. 96-24, Department of Information and Computer Science, University of California, Irvine.

[Franz & Ludwig 1991] M. Franz and S. Ludwig (1991); "Portability Redefined"; in *Proceedings of the Second International Modula-2 Conference*, Loughborough, England.

[Juice] M. Franz and T. Kistler; *Juice*; http://www.ics.uci.edu/~juice.

[Hansen 1974] G. J. Hansen (1974); *Adaptive Systems for the Dynamic Run-Time Optimization of Programs* (Doctoral Dissertation); Department of Computer Science, Carnegie-Mellon University.

[Ingalls 1971] D. Ingalls (1971); "The Execution Time Profile as a Programming Tool"; *Design and Optimization of Compilers*, Prentice-Hall.

[Oberon] Department of Information and Computer Science, University of California at Irvine; *Oberon Software Distribution*; http://www.ics.uci.edu/~oberon.

[Kistler & Franz 1997] T. Kistler and M. Franz (1997); "A Tree-Based Alternative to Java Byte-Codes"; *Proceedings of the International Workshop on Security and Efficiency Aspects of Java*, Eilat, Israel.

[Lindholm et al. 1996] T. Lindholm, F. Yellin, B. Joy, and K. Walrath (1996); *The Java Virtual Machine Specification*; Addison-Wesley.

[Inferno] Lucent Technologies Inc.; *Inferno*; http://plan9.bell-labs.com/inferno/.

[Nori et al. 1976] K. V. Nori, U. Amman, K. Jensen, H. H. Nägeli and Ch. Jacobi (1976); "Pascal-P Implementation Notes"; in D.W. Barron (Ed.); *Pascal: The Language and its Implementation*; Wiley, Chichester.

[Welch 1984] T. A. Welch (1984); "A Technique for High-Performance Data Compression"; *IEEE Computer*, 17:6, 8-19.

[Wirth & Gutknecht 1989] N. Wirth and J. Gutknecht (1989); "The Oberon System"; *Software-Practice and Experience*, 19:9, 857-893.

[Wirth & Gutknecht 1992] N. Wirth and J. Gutknecht (1992); *Project Oberon: The Design of an Operating System and Compiler*; Addison-Wesley.

[Wirth 1988] N. Wirth (1988); "The Programming Language Oberon"; *Software-Practice and Experience*, 18:7, 671-690.

Fetal Echocardiography Homepage

Krzysztof P. Wróblewski
Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine,
Philadelphia, USA
E-mail: krzyszto@mail.med.upenn.edu

Zhi Yun Tian
Division of Cardiology, Children's Hospital of Philadelphia, Philadelphia, USA
E-mail: tian@highresnmr.biophys.upenn.edu

Jack Rychik
Division of Cardiology , Children's Hospital of Philadelphia, Philadelphia, USA
E-mail: rychik@highresnmr.biophys.upenn.edu

Piotr M. Wróblewski
University of Pennsylvania School of Engineering and Applied Science, Philadelphia USA
E-mail: piotrw@eniac.seas.edu

INTRODUCTION

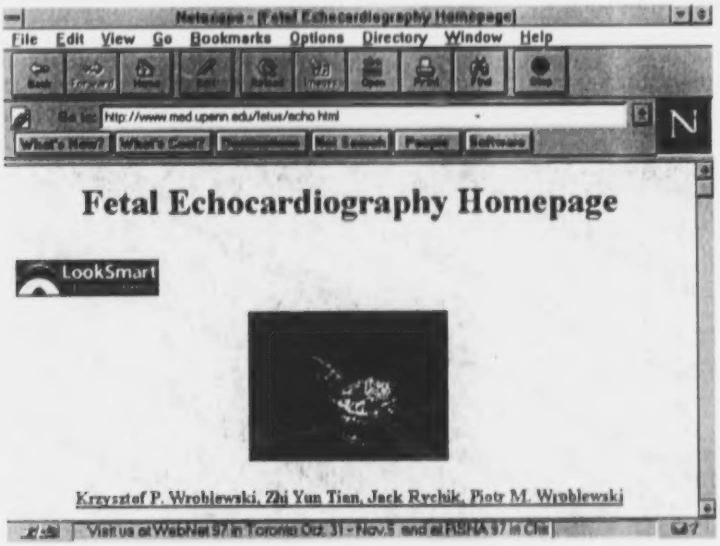
In the United States about 32,000 babies are born each year with heart defects. Until recently, little could be done for unborn babies suffering from anatomical abnormalities. Presently, improved fetal sonographic and sampling techniques, in conjunction with a better understanding of fetal pathophysiology, make therapy for the fetus an option [Adzick et al. 1996]. The most valuable and widely applied technique for evaluation of the human fetus is ultrasonography, which may be useful from the first week of gestation until the time just before birth. The ultrasonography of a heart is also called echocardiography. Every year a greater number of pregnant women are offered an ultrasound scan at approximately 18 weeks of pregnancy. The scan incorporates a detailed anatomical survey of the fetus, and if it includes at least a four chamber view of a heart, it is an excellent opportunity to detect most forms of congenital heart diseases [Huhta and Rotondo 1991]. While it appeared to have promise as a screening tool, it later became apparent that its potential benefit was limited by the experience of the physician or technologist performing the examination. In many cases obstetricians, or primary care physicians are not able to analyze images of the heart, and unless they are obvious many congenital heart abnormalities remain undetected.. The primary reason for this is lack of experience of the examiner. In addition, the orientation of the fetus presents a major problem. Unlike pediatric and adult cardiology in which standardized views of the heart are obtained, the fetus may present in a number of positions resulting in a myriad of orientations of the four-chamber and outflow tract views.

THE PROJECT

The goal for our project is to develop computer tools for effective teaching of medical personnel how to read and analyze echo data, and to support the process of detection of congenital heart abnormalities by non-cardiologists. We have already developed Fetal Echo Expert System [Tian and Wróblewski 1996], the artificial intelligence program capable of making diagnoses, and providing for early and appropriate detection of congenital abnormalities in fetus. The Fetal Echocardiography Homepage is an important part of the project. It was created in August 1996, and provides free information for medical students, residents, obstetricians and primary care physicians, including a library of congenital heart diseases It is located at the University of Pennsylvania School of Medicine World Wide Web server and it's URL is: http://www.med.upenn.edu/fetus/echo.html. The opening screen of the page is shown in Figure 1. The library of fetal echocardiograms consists of two units containing pictures of normal and diseased fetal heart. Both units include 2D echo, color Doppler, and M-mode pictures.

RESULTS

Since it's creation Fetal Echocardiography Homepage was accessed over 5000 times. A screenshot of the entire page has already been published in a textbook "Internet Resources for Cardiology" edited in Japan [PMSI Japan 1997]. The project was very well received by the American College of Cardiology [Wróblewski et al. 1997]. We are currently working on expanding the scope of this page. The number of letters we receive through the Homepage indicates that there is a great demand for a platform that can be used for an information exchange about different cases, and for distant consultations and diagnoses. The next step in the development will be creating possibility for uploading CINE loops containing important fragments of a study and have them reviewed by the experts and discussed publicly.



http://www.med.upenn.edu/fetus/echo.html

Fetal Echocardiography Homepage

Krzysztof P. Wroblewski, Zhi Yun Tian, Jack Rychik, Piotr M. Wroblewski

Visit us at WebNet 97 in Toronto Oct. 31 - Nov.5 and at RSNA 97 in Chi

Figure 1. The opening screen of Fetal Echocardiography Homepage

Fetal Echocardiography Homepage is written in HTML3 and Java Script. It contains links related to fetal cardiology sites and electronic journals where it already has been cited or referred to. The static images and animated pictures are stored in the widely accepted jpg and gif formats and can be retrieved and viewed with any World Wide Web browser with the Java extension, for example Netscape version 2.02 or higher.

REFERENCES

[Adzick et al. 1996] Adzick, N.S., Ballard, R., Crombleholme, T.M., Forouzan I., Graham, E., Morgan, M.A., Rose, N.C. (1996). Fetal Therapy Saves Lives. *Penn Today, 6, 9-10.*

[Huhta and Rotondo 1991] Huhta J.C., Rotondo K.M., (1991). Fetal Echocardiography. *Seminars in Roentgenology Vol XXVI, 5-11.*

[Tian and Wróblewski 1996] Tian, Z.Y., Wróblewski, P.M., Wróblewski K.P., (1996). Multimedia System for Fetal Echo Analysis. *Journal of the American College of Cardiology 1996, 9A.*

[PMSI Japan 1997] PMSI Japan Ltd. (1997). *Internet Resources for Cardiology* BIOMEDIS Tokyo Japan

[Wróblewski et al. 1997] Wróblewski, K.P., Tian, Z.Y., Wróblewski, P.M. (1997) Fetal Echocardiography Homepage *Journal of the American College of Cardiology - in press*

ACKNOWLEDGMENTS

We wish to thank Dr. Bogdan Kaluzewski for sending us fetal echo pictures and for the right to use them.

BUSINESS/ CORPORATE SESSIONS

International Tele-Education Experiment using CAI System on the World-Wide Web

Yasuhisa KATO*, Jun HIRAI*, Yoshimi FUKUHARA*, Mitsuru Yamada**
*NTT Information and Communication Systems Laboratories
** KDD
E-mail: yasu, hirai, fukuhara@isl.ntt.co.jp, mu-yamada@kdd.co.jp

1. Introduction

This paper describes the results of the international trial for CALAT(*1) [Nakabayashi et al. 1995] between Keio Schools in Japan and USA. The objective of this trial is to evaluate the usability of CALAT, especially in case of using high-speed network such as ATM.

(*1)CALAT: Computer Aided Learning and Authoring environment for Tele-education. This is an adaptive individual tutoring system, which is integrated in the distributed hypermedia environment of the World-Wide Web (WWW). Users can learn materials on remote servers using the WWW browser through the network.

2. Experiment

The focus of CALAT testing was the quick response capability when using the high-speed international ATM link. We measured the basic response performance in accessing several files on CALAT servers located in the U.S. and Japan sites. And under this condition, after the use of CALAT by Keio School students, they answered the questionnaires concerning usability and capability of CALAT.

(1) Response time measurement
We measured the response time between the request of the client and the playback of the appropriate media in the client. The measurements were performed under the same conditions except VP shaping. The response time was the average of two trials. TCP window size was the maximum 64 KB on the contrary of the default 8 KB, because the response time strongly depended on TCP window size, and 64 KB was the best condition.

(2) User questionnaire
After about an hour usage of CALAT, students answered the quality of sound, the usability, the response speed, capability, and so on.

3. Results

(1) Response time measurement
1. As for only one user (client), 1.5 Mbps was enough and 10 Mbps was excess.
2. When the learning material had smaller files, that were about 200 KB, it was reasonable to use the transpacific international remote server.
3. Using a 85 KB file, the response time of the remote server was as same as the local server. But using a 179 KB file, it was 1.5 times. Moreover, using 4 MB it was seventeen times.
4. It took more than two minutes to transmit 4 MB movie file even in 10 Mbps shaping.

(2) User questionnaire
1. Sound quality, GUI were good. Response speed was fair. The interest and usability of CALAT were very big.
2. In each experiment, about 30 students at most accessed to 3 CALAT servers simultaneously. As each student accessed to a server he/she liked, about ten students in the average were getting data from each server. From the questionnaires, the response speed was fair. The capability of the transpacific server was proved.
3. A lot of students had interests with their familiar topics.

4. Conclusion

1. 1.5 Mbps international connection was enough for the usual learning materials. It depends on TCP protocol. TCP Gateway [Hasegawa 95] will be needed using more than 1.5 Mbps line.
2. It took a long time to transfer a huge video data. 10 Mbps bandwidth was insufficient. It also depends on TCP protocol.
3. There was a big concern and demand in such a learning material using the WWW.

5. Future work

1. Some streaming techniques, such as a streaming audio, video, and animation are needed for reducing the silence of the system.
2. We should make much variety of learning materials. Some materials are interesting and others are boring. It depends on the individual. We should provide much variety of materials.

Acknowledgements

This research was supported by Multimedia Application Project [Fujii 97] which AT&T, KDD, and NTT joined. We appreciate the support of Keio Academy, especially Keio Academy of New York, Keio Chutobu Junior High School, Keio Yochisya Elementary School, and Keio Shonan Fujisawa Junior and Senior High School.

References

[Fujii 97] Hiroyuki Fujii, Yohnosuke Harada, Shigeaki Tanimoto, Seiji Kihara, Hirohisa Miyashiro, Kanji Hokamura, Mitsuru Yamada, Toshihiko Kato, Linda Galasso, "An Experimental Study of Multimedia Applications over International ATM Networks," APCC97, 1997

[Hasegawa 95] Teruyuki Hasegawa, Toru Hasegawa, T. Kato, and K. Suzuki, "Implementation and Performance Evaluation of TCP Gateway for LAN Interconnection through Wide Area ATM Network," IEICE Trans. Comm., vol.J79-B-I, no.5, pp.262-270, 1995 (in Japanese)

[Nakabayashi et al. 1995] Kiyoshi Nakabayashi, Yoshimasa Koike, Mina Maruyama, Hirofumi Touhei, Satomi Ishiuchi and Yoshimi Fukuhara, "A Distributed Intelligent CAI System on The World-Wide Web," ICCE 95, pp. 214-221, 1995

Remote Lecture between USA and Japan over the ATM Network

Naoko Takahashi, Hisasumi Tsuchida, Yoshimi Fukuhara
NTT Information and Communication Systems Laboratories, Japan,
{naoko, tsuchida, fukuhara}@isl.ntt.co.jp

Yoshiyasu Takefuji
Professor, KEIO University, Japan
takefuji@sfc.KEIO.ac.jp

Raymond K.
Case Western Reserve University , Cleveland, Ohio
rkn@po.cwru.edu

Introduction

At the end of 1996, several experimental remote lectures has enforced between KEIO University(Japan) and Case Western Reserve University(U.S.) as an multimedia application on international connection experiment for ATM between Japan and U.S.[Fujii,1997].

In this paper we describe the results of this remote lecture and discuss the requirements for sending remote lectures over an ATM network.

Experiment

We held experimental remote lectures in which Shonan Fujisawa Campus (SFC) of KEIO University in Japan was connected with Case Western Reserve University (CWRU) in the U.S.. Dr. Takefuji in a laboratory at SFC as a teacher taught the total number of more than ten information processing graduate students in a remote lecture room at CWRU.

We deployed workstations on SFC and CWRU, and connected between the each workstation over the ATM network. The connection between the each site was the maximum bandwidth of 10 Mbps. And the workstation on the teacher site also connected to the Internet.

We used the 'Communique!' as videoconferencing application so that the teacher and students could see each other and the students could ask questions. The teacher used the WWW as one source of teaching materials. And to show the students various materials (text on paper, a PDA, or a notebook computer), the teacher used an overhead camera to capture text and pictures.

During the lecture, the teacher was operating all applications and switch, and he was controlling the direction of the camera on the students' side himself. On the students' side, one operator was controlling a sound volume and a layout of applications windows.

We evaluated the effectiveness of the four lectures by measuring the ATM cell-transfer rates, by administering a questionnaires to the participants and by the interviewing the teacher.

As a trial, we also connected between the same places via the Internet and the video image was sent by 'Communique!'. And then, we evaluated the quality of 'Communique!' over an ATM network by comparing with that on the Internet.

Results

In this section, we discuss requirements for remote lectures over the ATM network.

(1) Quality of the video image

When we used 'Communique!', the average ATM throughput was 1300 kbps, and the average frame rate was 22 fps. And the average throughput over the Internet was 900 kbps, and the average frame rate was 15 fps. On this questionnaires, eight of the ten students considered the video image to be "average" or "smooth".

(2) Quality of audio

Nine of the ten students described the audio delay as "could not notice" or "very short". Six students described the audio as "clear" while four said it was "a little unclear". During the Q&A., severe echoes arose, making it hard to establish bi-directional communication. Nine of the students complained about this problem.

When we used 'Communique!' over the Internet, the audio frequently interrupted and was less clear than over the ATM network.

(3) Teaching materials

The workstation on the teacher's site was connected to the Internet so that the teacher provided such information on the Internet as the Web for teaching materials. the teacher could use various media and the newest information , and give more effective lectures to the students.

However, when we used a shared application function to display the same Netscape Browser window at each site, it took more than twice as long as without this function. And there were no appropriate (and easy) way to present such things as PDA and notebook computer displays, we could only use the overhead camera.

(4)Time difference

These lectures started at seven o'clock in the evening at Cleveland(students' side) and nine o'clock in the morning at Kanagawa(teacher's side) side because there is a fourteen-hour time difference between Cleveland and Kanagawa. It led the reduction of the number of participants that these lectures started too late for the students. It is difficult to perform remote lectures periodically. Rather, it is more effective to take remote lectures as complements to regular off-line lecture courses.

(5) Usability

The teacher could operate the direction of the camera on the students' site to check the condition of the students. But he couldn't see the display on the students' site and it is difficult for the teacher to communicate with the operators, so that the view point of the students sometimes wasn't different from the view point where the teacher wanted for the students to look at.

(6) Remote lectures in the future

On this questionnaire, we got the following opinion about remote lectures in the future.

In the recent future, Tele-education will become very important. It is necessary to develop a remote lecturing system which is economical and simple that anyone can easily use. And anyone will be able to attend various lectures of universities at home and acquire a degree from home.

Conclusions

(1)On the quality of the video image and the audio, it is difficult to hold remote lectures over the Internet. On the other hand, a 10-Mbps international ATM connection provides adequate performance for real-time remote lecturing.

(2)The audio channel is more important than the video channel for remote lectures.

(3)It is necessary for teachers to show students various media for the teaching materials. Specifically, it is necessary to display all pages which teachers get from the WWW in real time.

(4)A seamless environment must be provided in which teachers and students forget about the distance separating them. That is the teacher should know the state of students and be able to relate to them as in a real classroom environment.

Acknowledgements

This research was supported by Multimedia Application Project [Fujii 97] which AT&T, KDD, and NTT joined.

References

[Nakabayashi, 1997] Kiyoshi NAKABAYASHI, Mina MARAYAMA, Yoshyimasa KOIKE, Yasuhisa KATO, Hirofumi TOUHEI, and Yoshimi FUKUHARA(1997). Architecture of an Intelligent Tutoring System on the WWW. Artificial Intelligence in Education. 1997. Knowledge and Media in Learning Systems.

[Fujii 1997] Hiroyuki Fujii, Yohnosuke Harada, Shigeaki Tanimoto, Seiji Kihara, Hirohisa Miyashiro, Linda Galasso, "Multimedia Application Project over International ATM Networks," APCC97, 1997

Working Together While Being S-e-p-a-r-a-t-e-d

G. van den Broek, Philips Research IST, Netherlands (vdbroek@natlab.reserach/philips.com)
F.-J. Stewing, SNI C-lab, Germany (stewing@c-lab.de)

Abstract: Within industry collaboration across time and space is becoming more and more important. This collaboration involves several disciplines (engineering, marketing, service and support) and sometimes also several companies. Because there are already a lot of tools on the market, and they are evolving very fast, the MATES project approaches this problem from the user view. A reference model for services has been defined, and currently available tools are investigated how they fit within the reference model. Some weak spots have been identified and solutions for them will be implemented. An integration framework for asynchronous as well as synchronous communication between tools will be used to integrate a number of existing tools and tools developed within the project. Special emphasis will be given to the integration at the users desktop (UNIX as well as Windows/NT) where the information and tools will be presented to the user within the context of a project support environment. Already during the project the partners have been using available tools to get experience with this new technology.

Introduction

Industry today is confronted with an increase of the complexity of its products and also by a need to reduce time-to-market. This often leads to situations where not all the expertise or resources can be found in one location or inside one company. New models for product creation processes, such as co-design, interaction with product marketing and co-makership, are being introduced to companies. Also the restructuring of companies into smaller more independently operating units increases the need for collaboration across sites and sometimes across companies. The manner of how distributed working is exploited within an organization requires specific attention to the aspects of globalisation of the development process, of project management, of data management and of the engineering tasks themselves. The generic problem can be described as: *supporting the process of getting consensus about topics to solve and enable sharing of knowledge between persons separated in time or space*.

The MATES project, which is an ESPRIT funded project (EP 20.598), is aiming to support such collaborative working by offering multimedia-assisted distributed tele-engineering services that can be used to construct Distributed Engineering Environments or Interactive Remote Maintenance Environments. Most of these services are of a generic nature, i.e., they might be used or shared with other disciplines as for example marketing, product management.

Collaboration Services

Services which should support collaborative working are called collaboration services. The technologies related to these services are evolving quite fast. To be able to cope with this evolution a reference model has been defined within MATES. This reference model enables required services to be specified and discussed independent of the final implementation. It also provides a tool to assess the current situation, to propose improvements, to formulate an implementation plan with priorities, and to assess products to make the required services available.

From a functional point of view the Collaboration Services can be categorized in:

- Communication Services

  The communication services cover the exchange of data (in all kind of formats) between a defined list of persons. This means that data will be sent either synchronously or asynchronously to defined sets of recipients. The exchange can also be categorized as having a low structural complexity and can vary from a more passive nature to an active nature. These services include:

  - e-mail in combination with MIME attachments to be transfer any kind of file format);
  - audio and video conferencing;
  - remote presentations (in combination with audio and optionally video);

  - electronic white board (in combination with audio and optionally video);
  - application sharing (in combination with audio and optionally video).

- Cooperation Services

  The cooperation services enable users to make their (intermediate) results available to others and to participate in discussion forums. It is often required to control the group of persons having access to the data involved. Traditionally this kind of services is often called groupware. These services must provide support for:

  - a project and public archives to upload files/documents, maintain indices on these files and documents, and support searching across documents a.For project based archive access control should be possible;
  - public and project- or team-wide discussion forums, with facilities to set-up forums, manage access lists and moderate the discussions;
  - calendar and scheduling facilities;
  - services to maintain project data such as membership, address lists, location of archives.

- Coordination Services

  The way a (product creation) project has been organized, the working procedures and the allocation of responsibilities, highly influence the way people will collaborate. The coordination services allow people belonging to a team to work on the same set of files in an organized and controlled way. In other words the services should support the access to the work in progress. The services required are:

  - support for a shared project workspace where project members can store their data and access data from co-workers;
  - access to functions of existing applications (especially managerial tools) which are already in use in the organization, such as a configuration management system or an EDM system, workflow system, process support system, change control system,...

Collaboration Services and Process Models

The organizational processes highly influences the of collaboration between parts of an organization. To get some feeling of these aspects some models are described below.

In the **Supplier - Customer** model one see a one to many relation with a heavy information (e.g. brochures, catalogue information, mailing etc.) flow from the supplier to the customer and a limited flow in the opposite direction (orders, product feed-back). In an electronic world this could be accommodated by WWW accessible document stores and e-mail. The document stores could divided in public part and a protected part for registered customers.It would also be useful for customers to enter problem reports.

In a **Sub-contracting** situation the relation is more balanced and the activities to be supported are: getting consensus and having access to a well-defined set of data. The consensus making process can be supported with e-mail, threaded discussion forums and audio/video conferencing, while the access to the product data and documents would require a project documents store. This document store can also be the base for reviewing purposes while a calendar and scheduling tool could give some support in defining and administrating dead-lines.

In a **Branch development** (independent development of variants) situation the collaboration is oriented to exchange knowledge or inform each other on a peer-to-peer relation. This can be supported with e-mail and threaded discussion forums, while the access to the product data and documents would require multiple project documents stores (one for each partner).

In the **Co-design** models collaboration the goal is to achieve consensus on certain topics or approach. The consensus making process can be supported with e-mail, threaded discussion forums and audio/video conferencing, while the access to the product data and documents would require multiple project documents stores. A shared project work-space might be useful to make work-in-progress data available to each other. A calendar and scheduling tool could give some support in defining and administrating dead-lines.

A more complex model is based on the process of **Co-makership** where a number of team are supposed to work together on single product. To make this possible one needs to support a coordinated way working between the members of those teams. In addition to services mentioned above for co-design one should also support some kind of shared project workspace and probably a number of tools from a managerial kind of nature (such as; configuration management, problem tracking, workflow and process support) accessible for all members of such a distributed project.

From the discussion above one can derive some indication for intensity of collaboration and its importance in the context of distributed engineering (see also "Figure 1: Collaboration Services and Process Models").

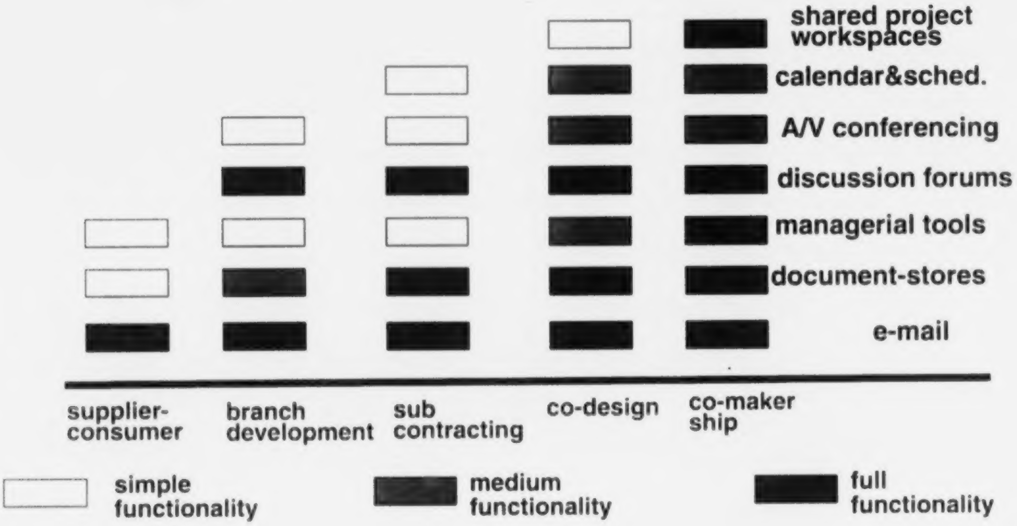


Figure 1: Collaboration Services and Process Models

The MATES Approach

A distributed working environment should enable knowledge intensive development projects to perform efficiently independent of the physical location of their participants. This mean that the process of getting consensus should be supported and secure and safe access to project and product data should be available. An important aspect is to support the engineering processes and procedures by adding distributed access to existing applications, which might already be in use, and their data.

Despite the existence, for some time, of technology relevant for MATES the awareness of its existence still needs to be increased and its introduction for wide usage needs to be simplified. An additional complication is the rapid evolution of available technology. However, there is no single, universal solution ('no silver bullet') which covers all needs for distributed working. Depending on the specific needs and priorities of a given organization a set of selected services must be offered by selecting existing tools (as far as possible). This set of selected tools must be integrated by constructing a distributed environment by the way open framework. Such a framework based approach makes it possible to take advantage of existing applications and with to cope with the rapid evolution of new applications. Solutions must be open and configurable, depending on the individual distributed engineering application situations. Combinations of technologies like WWW, CORBA, OLE, etc. are used as basis for the MATES framework to enable the required flexibility. "Figure 2: Mates Architecture" gives a simplified view of the Mates architecture. The communication, cooperation and coordination tools are integrated into a project support environment. Individual components are distributed across the available communication infrastructure (which might be a wide area network). The tools can exchange control messages and/or application data using the MATES framework. The solutions offered by the MATES project must also be able to work in a wide area network which might include the public Internet. Obstacles related to the communication network for bringing MATES-like technology into practical use are: the available network bandwidth (e.g. the unpredictable performance of the public Internet) and security issues such as: firewalls, which protect the company networks, secure exchange of data across public Internets, and authentication to check the identification of the user. This issues are not tackled by the MATES project itself. We are rely-ing on tools on the market to solve this.

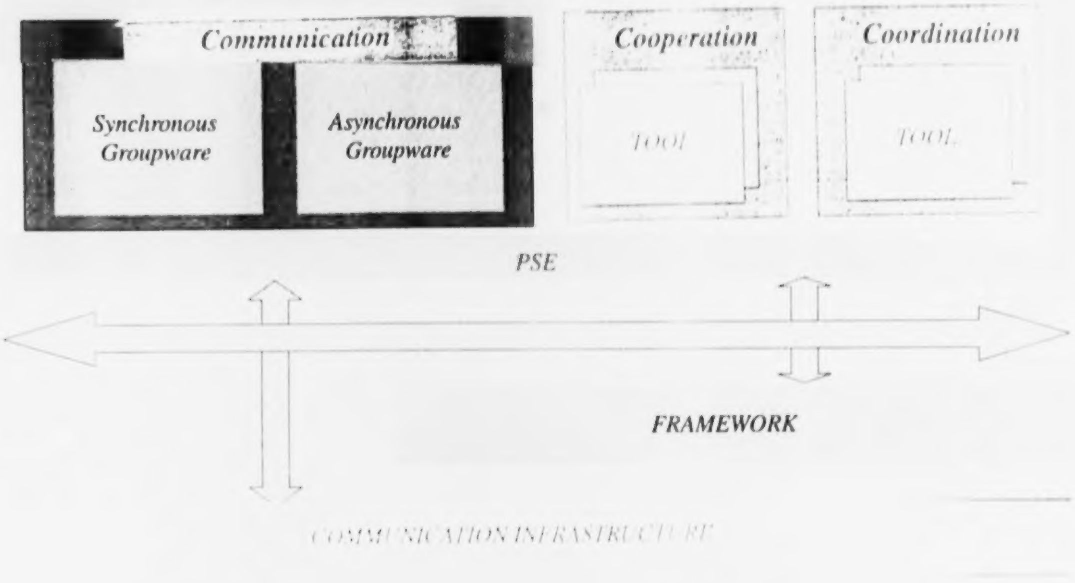


Figure 2: Mates Architecture

As observed before a number of applications are already available, either as public domain tool or as commercial product, to accommodate part of the requirements. This holds especially for the communication and cooperation categories. The solution for the coordination category depends highly on the working procedures of the teams involved (process model) and the application domain to be supported (e.g. software engineering, IC design). The added value of the MATES project will be the integration of existing applications in a framework which supports the distribution of data and applications and making these data and applications easily accessible from the engineers desktop (UNIX and Windows/NT) within the project context.

Table 1 below gives the reference model presented earlier filled in with components for a proposed solution. Tools in *italic* are existing tools, while tools in **bold** are part of MATES developments.

Conferencing and presentation applications have been developed by CDT. These applications are based on Internet multicasting facilities and consist of audio and video conferencing, whiteboard, and a Web based presentation application. An application to record and play-back a multicast based conference has been developed as well. These tools together are called m*environment.

An application sharing facility for the X environment is available from SNI/ASM. They have also developed an application sharing agent which will use MBone and will be integrated into the conferencing solution of CDT. JointX is an existing version of this application.

A CORBA based integration platform (LiP) has been developed by SNI/C-lab. This integration framework offers also a workflow component. SNI/C-lab will also implement the Project Shell which is the main user interface and offers easy access to all project relevant data and supports the administration of the project data.

The University of Madrid has developed a CORBA interface for the popular CVS configuration management stem.

BEST COPY AVAILABLE

	Communication	Cooperation	Coordination
Activities	addressed exchange of any type (text, audio, video, documents) of data	access to released results and discussion forums with optional access control	controlled access to work in progress
Services	e-mail with attachments, audio and video conferencing, remote presentations, remote application sharing	threaded discussion forums project and public archives search facilities, membership administration	shared project workspace, access to existing applications (EDM, problem tracking, workflow, process support)
Standards	H.32x&T.120, MBone, SRM, RTP, SRRTP, vCalendar, vCard		
	SMTP, TCP/IP, IMAP, ICAP, http, html, MIME, JAVA, CORBA, IIOP		
Tools	H.32x and T.120 based conferencing tools *(PictureTel, Proshare, neT.120, Netmeeting)* **MBone tools: JointX, m*environment**	Commercial forums Commercial doc. stores *(Domino/Notes, AltaVista Forum, LiveLink, Netscape Suitespot)* **ProjectShell**	*BSCW*, **LiP** **Integrations of:** *Continuus CM/PT* *CVS*
	Integration on the desktop and with Web-browser		

Table 1: The MATES Solution Within the Reference Model

Status of the MATES project

The MATES project started on February 1st, 1996 and expects to end at 1st of July 1998. In the MATES project, Dassault Electronique (Paris, France), Philips (Eindhoven, the Netherlands), and Telefonica I+D (Madrid, Spain) are participating as user organizations which provide requirements and take part in the evaluation tasks. Center for Distance-spanning Technology (Lulea, Sweden), University of Madrid (Madrid, Spain) and Siemens Nixdorf Informationssysteme (Berlin and Paderborn, Germany) will contribute applications and technology.

A consortium like this, working on the topics of distributed engineering and with such a geographical distribution, is of course challenged to use at least part of its proposed solutions to support its own work. This challenge has been taken up within the project we are MIME based e-mail is used, a Web-server which offers threaded discussion forums (using HyperNews) as well as a project workspace (using BSCW) and a document archive for accepted deliverables have been installed. Synchronous communication using the public Internet has been used by CDT for their weekly project meetings. For multi-point video-conferencing between the different partner we use PC based video-conferencing on ISDN.

The results from MATES will be:

- a (refined) reference model of services for distributed collaborative engineering. Such a reference model can be used for assessing an organization and it communication infra-structure, for planning improvements and to evaluate tools.
- guidelines to construct distributed collaborative engineering environment;

- more clear information on computer and network requirements and its limitations;
- a framework and integration technology to construct distributed collaborative engineering environments;
- an application (ProjectShell) which is configurable to a kind of "virtual project room";
- building blocks which can be used as part of such distributed collaborative engineering environments.

These results of MATES will be used to finally construct an interactive remote maintenance support system and a distributed engineering environment. These two integrated environments represent the MATES Evaluation Pilots and should demonstrate the "better-than-being-there environment to support working together while being s-e-p-a-r-a-t-e-d.

References

In the following references are given to (published) papers, articles and slides on MATES, mainly being available on the MATES Webserver ***http//mates.cdt.luth.se:***

[Broek 1996] v.d. Broek, G.(1996). *"Aspects of Distributed Engineering"*, 1st GEN Workshop, Paderborn, Germany

[Broek, Faux and Stewing 1997] v.d. Broek, G., Faux, I. and Stewing, F-J. (1997). "Linking EDM/PDM Systems With Support for Collaborative Work: The MATES Approach", PDT Days '97, Sophia Antipolis, France

[Broek and Stewing 1997] v.d. Broek, G., Stewing, F-J. (1997). "MATES: MultiMedia Assisted Distributed Tele-Engineering Services", ICED'97, Tampere, Finland, August 1997

[Parnes, Schefstroem and Synne 1995], Parnes, P., Schefstroem, D., & Synne, K. (1995). *"WebDesk: The Tele-conferencing Services of MATES"*, SMAC-Workshop, Sligo, Ireland

[Parnes, Schefstroem and Synne 1996], Parnes, P., Schefstroem, D., & Synne, K. (1996).*"WebDesk - Collaboration Support in MATES"*, 1st GEN Workshop, Paderborn, Germany,

[Parnes, Schefstroem and Synne 1997], Parnes, P., Schefstroem, D., & Synne, K. (1997). *"The mWeb Presentation Framework"*, 2nd GEN Workshop, Antwerp, Belgium

[Parnes, Schefstroem and Synne 1997a], Parnes, P., Schefstroem, D., & Synne, K. (1997). *"The WebDesk Presentation Framework", INET'97*

[Parnes, Schefstroem and Synne 1997b], Parnes, P., Schefstroem, D., & Synne, K. (1997). *"mood: the Multicast Media-on-Demand system", NOSSDAV '97*

[Stewing 1996] Stewing, F-J. (1996). *"Working Report Engineering in the Virtual Enterprise"*, 1st GEN Workshop, Paderborn, Germany, February 1996 Stewing, C-LAB, "MATES: Multi-media Assisted Distributed Tele-engineering Services", Colloqium 10th C-LAB Anniversary, Paderborn, Germany

[Stewing 1996a] Stewing, F-J. (1996). *"Enterprise Integration Through the Global Engineering Network, GEN, and MATES"*, CALS Europe 1996, Paris, France

[Stewing 1996b Stewing, F-J. (1996). *"IT and Infrastructures for Sustainable Development: Contributions Emerging from GEN and MATES"*, Advanced Course on Sustainable Industrial Production, Venice, Italy

[Stewing 1996c Stewing, F-J. (1996). *"Tooling and Procedures in MATES"*, 1st GENIAL Workshop, Paderborn, Germany

[Stewing 1996d Stewing, F-J. (1996). *"GEN, GENIAL and MATES: Turning Engineering Knowledge into an Accessible Corporate Asset"*, EITC '96, Brussels, Belgium

Acknowledgements

The work performed in the MATES project is part-funded by the ESPRIT programme under contract EP 20.589. In the MATES project, Dassault Electronique, Philips, and Telefonica I+D are participating as user organizations that state their requirements and take part in the evaluation tasks. CDT, UPM and SNI will contribute with applications and technology.

The presented ideas are the collective views and work of the MATES project management team (R. Campo (Dassault Electronique), C. Bathe, D. Schefstroem (CDT), L.F. Solorzano (Telefonica), J.C. Yelmo Garcia (UPM)), and all other MATES project members which couldn't all be mentioned here, but to which thanks needs to be expressed at this place.

FULL PAPERS

Shared User Behavior on the World Wide Web

Ghaleb Abdulla, Edward A. Fox and Marc Abrams
Network Research Group, Computer Science Department, Virginia Tech
Blacksburg, VA 24061-0106
{abdulla, fox, abrams}@vt.edu

Abstract: Studying accesses to Web servers from different user communities helps identify similarities and differences in user access patterns. In this paper we identify invariants that hold across a collection of ten traces representing traffic seen by proxy servers. The traces were collected from university, high school, governmental, industry, and online service provider environments, with request rates that range from a few accesses to thousands of accesses per hour. In most of the workloads a small portion of the clients are responsible for most of the accesses. In addition most of the accesses go to a small set of servers. By doing a longitudinal study on the collected data we noticed that the identified invariants do not change over a year period. However, the percentage of script generated documents, is increasing .

Introduction

In recent years the World Wide Web (WWW or Web) has grown rapidly as a dissemination tool for different kinds of information resources. Frequently, the Web is used for deployment of educational and commercial material. Educators are using the Web to post course notes, syllabi, homework assignments, and even exams and quizzes. Companies are using the Web for advertising, publicity, and to sell products.

The dynamics of Web traffic are not well understood. There are several differences between the Web and other types of network traffic. Those differences emerge from the HTTP protocol used and Web users' behavior. With respect to the HTTP protocol, clicking on hyperlinks that are part of HTML pages generates traffic and, as a result, a new HTML page or an image is displayed. HTML pages contain formatted text and graphics. Sometimes links in HTML pages lead to other types of media, such as video or audio. In contrast, traditional network traffic has formatted or unformatted text, and rarely uses graphics, video or audio. With respect to users, the low level of expertise required to navigate with a Web browser has resulted in a large and diverse user population. Therefore, it is reasonable to assume that Web users behave differently from those who use other network resources. The status of Web servers and network connections and how fast they can respond is a factor that affects future accesses by users.

In this paper we examine ten traces that were collected from university, high school, governmental, industry, and online service provider environments, with request rates that range from a few accesses to thousands of accesses per hour. We analyze the traces in order to understand the way users interact with the Web and to explore if users with different backgrounds display different behavior when using the Web. We look for invariants that hold across the traces.

We examined the collected traces to find out if there are similarities between accesses from educational institutions versus accesses from industry, government, or home. We study accesses made by a group of users who either share the same workplace (and they are potential users of a *proxy* server if available) or use a *proxy* server. A proxy is a server that can act as a cache and a gateway. It can send requests for Web documents as well as serve Web documents from its cache. A company might not have individual PCs on the Internet for security, Yet the PCs are given Web access by using a gateway or a proxy. For a group of clients a proxy looks like a Web server and for a Web server it looks like a client. The browsers on the client side can be configured to point to the proxy so that any access from the client goes first to the proxy. There is a growing interest in proxies for caching, Web TV, and cellular phones, hence it is important to study and characterize accesses to proxies.

Related Work

Several studies have characterized client workloads [Crovella & Besravros 1996] and server workloads [Arlitt & Williamson 1996]. However, we have found no published study to characterize proxy workloads. This is due to the difficulty of collecting proxy log files from different sources and the privacy issues in information contained in such logs.

Arlitt and Williamson used six different server log files to characterize accesses; they identified ten different invariants for Web server workloads. The invariants in the study were used to identify two strategies for cache design and to determine the bounds on performance improvement due to each strategy.

In [Cunha & Besravros 1995] and [Crovella & Besravros 1996] the data was collected from a group of clients accessing the Web. The authors in [Cunha & Besravros 1995] showed that many characteristics of the WWW can be modeled using power-law distributions such as the *Pareto* distribution.

In this paper we characterize the traffic seen by a caching-proxy by identifying a set of invariants that hold true accross the examined traces. We compare traffic from educational, governmental, commercial and home users to see if the traffic generated differs between communities.

Objectives

This study is part of a comprehensive effort to characterize proxy workloads and test if invariant properties exist that hold across many proxy workloads from different communities. We also test if some of the identified invariant properties hold true over a year period.

Workloads Studied

Workload	Period	Accesses	Bytes(MB)
DEC1	9/3/96	1304565	11206.99
EDC2	9/19/96	1293147	10889.42
BU(G)	11/29/94-2/27/95	52901	293.99
BU(U)	1/27/95-2/22/95	414350	1201.18
Korea	9/2/95-9/26/95	1681963	21941.65
VT-Lib	9/19/96-11/20/96	127853	589.21
VT-CS	1/1/96-11/18/96	570385	3491.74
VT-Han	7/12/96-11/20/96	440345	2577.67
AUB	10/21/96-10/22/96	19259	109.52
AOL	12/96(few minutes)	883082	6017.88

Table 1: Summary of workloads used.

Tab. 1 summarizes the workloads used in this study, showing dates of collection. Collection procedures differ between workloads. Workloads from Virginia Tech, Computer Science (VT-CS), Hancock Hall (VT-Han), and the main campus Library (VT-Lib) as well as Auburn high school (AUB) in Virginia, were collected using a tool called *httpfilt* [Abrams & Williams 1996]. Before analyzing the data we used a filter to exclude accesses to local servers. This way we only examine accesses to remote servers; that is what a proxy would see. The Digital workloads DEC1 and DEC2 were collected using a modified version of the 1.0.beta17 squid proxy [Digital 1996]. The modified proxy was installed on two machines that act as Web proxies for Digital's internal network. The Boston University log files BU (G) for graduate students and BU (U) for undergraduates were collected by modifying a version of Mosaic, which was popular at the time of collection, to record certain information for each client [Cunha & Besravros 1995]. The America On Line (AOL) trace also was collected using a proxy server; however their log file only contained a list of URLs accessed by users and it did not have clients, sizes, or timing information for privacy reasons. Using the available URLs and a locally developed software tool called *WebJamma* we replayed all AOL accesses through a modified Harvest server and generated a new log file which includes file size information [Wooster and Abrams 97]. The Korea log file was collected using a proxy server installed as the gateway to South Korea. Therefore, the logs contain all trans-Pacific traffic.

Proxy Workload Invariants

Tab. 2 lists invariants that hold true across the workloads studied. These invariants are discussed in detail in [Abdulla et al. 1997]. We follow closely the work done in [Arlitt & Williamson 1996] in establishing the invariants. However, since our workloads are for a different class of HTTP traffic, namely traffic seen at a proxy server, we compare it with the identified invariants for servers.

Number	Name	Description
1	Median file size	approximately 2KB
2	Mean file size	less than 27KB
3	File types (accesses)	90%-98% of accessed files are of type graphics, HTML, or CGI-map
4	File types (bytes)	most bytes accessed are of type graphics
5	% of accesses to unique servers	less than 12%
6	% of servers referenced one time	less than 5%
7	Accesses concentration (servers)	25% of the servers get 80%-95% of the total accesses
8	Bytes concentration (servers)	90% of the bytes accessed are from 25% of the total number of servers
9	Success Rate	88%-99%

Table 2: Invariants for the workloads in Tab. 1.

The median size in all workloads is very close to 2K; the minimum among the medians for tested workloads is 1938 and the maximum is 2658 bytes. The mean file size ranges between 7K and 27K. These two findings are listed as invariants 1 and 2 in Tab. 2. Our findings here are consistent with the findings from the server workload study. Since accesses to proxy servers represent samples from thousands of servers around the world, there is evidence that invariants 1 and 2 may apply to many other servers.

Graphics files are the most frequently accessed file type in all workloads. To reach the level of 90% of the total accesses, though, in contrast to the server invariants, CGI-bin and image-maps (which we refer to as CGI-map), files also must be considered. In our traces HTML, graphics and CGI-map files represent 90%-98% of the accessed files. Invariants 3 and 4 in the table list these two facts. By examining the file types and comparing times of collection we concluded that the percentage of dynamically generated documents, such as CGI-map has increased. We also noticed that types such as video, audio, postscript and Adobe's portable document format are increasing in percentage and the bytes transferred for such types are significant.

To find out the percentage of accesses to unique servers, we sorted accessed servers by name and counted every server once. We found out that accesses to unique servers are less than 12% in all workloads, so approximately 90% of the repeated accesses are to the same set of servers. This is invariant 5 in Tab. 2. As we expected, the percentage of servers accessed only once is very small, less than 5%; this is invariant 6 in Tab. 2. We also did similar tests to accessed URLs; although the DEC traces behaved in a different way, the rest of the traces showed that a small percentage of the accesses go to unique URLs, and URLs accessed one time only represent a very small percentage. The previous results suggests that the locality of reference is very high for all workloads and that caching should be effective. To check this assumption we use a simulation to check the hit rate and the weighted hit rate for the workloads; see [Abdulla et al. 1997].

To examine the distribution for clients' accesses and the distribution of accesses to servers and URLs, we plot percentage of accesses vs. percentage of servers, URLs, and clients. Fig. 1 shows that 25% (represented by the vertical line in the figure) of the servers get 80%-90% of the accesses while the other 10%-20% get the rest of the accesses. This is invariant 7 in Tab. 2.

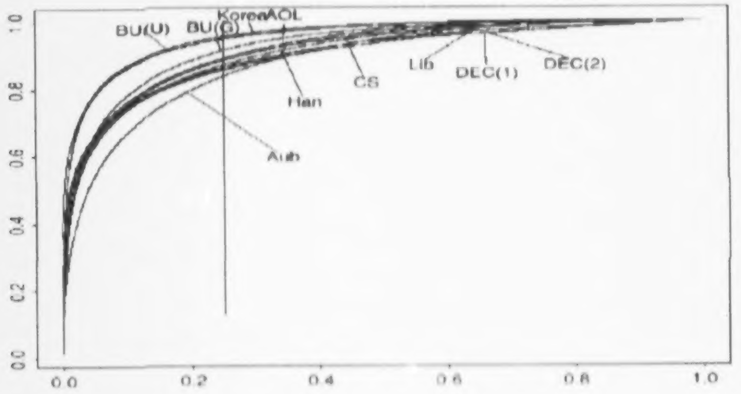


Figure 1: Servers concentration of accesses.

By examining Fig. 2 we see that in all workloads except DEC 85%-95% of the accesses go to 25% (the vertical line in Fig. 2 shows the 25% mark) of all URLs. All workloads except DEC have similar graphs and access distributions.

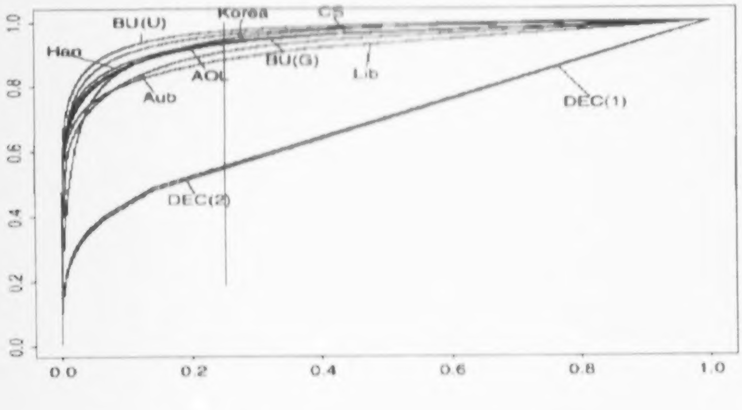


Figure 2: URLs concentration of accesses.

We examined client access distributions to see if we can find invariants across workloads. The percentage of clients versus percentage of accesses made by those clients is plotted in Fig. 3. Except for the Boston workloads, in all other workloads 50% (see vertical line in figure) of the clients are responsible for 80%-95% of the accesses. This could be true because of the existence of multi-user machines where users can login and run multiple instances of the network browser. Two cases that represent the extreme points in the graph, Computer Science at Virginia Tech (VT-CS), and Boston University BU (G) are collected from similar environments, that is, from computer science departments and graduate students.

However in reality the behavior is completely different. The BU (G) curve is almost linear and approximately 50% of the clients are responsible for 50% of the accesses. On the other hand, 50% of the clients in VT-CS workload are responsible for more than 95% of the accesses. One reason for the difference is that in VT-CS most of the accesses come from one multi-user machine. Other machines are lab machines used by various students or office machines. In the BU (G) workload there are only five workstations; however in the VT-CS workload we have over thirty clients most of which are multi-user machines.

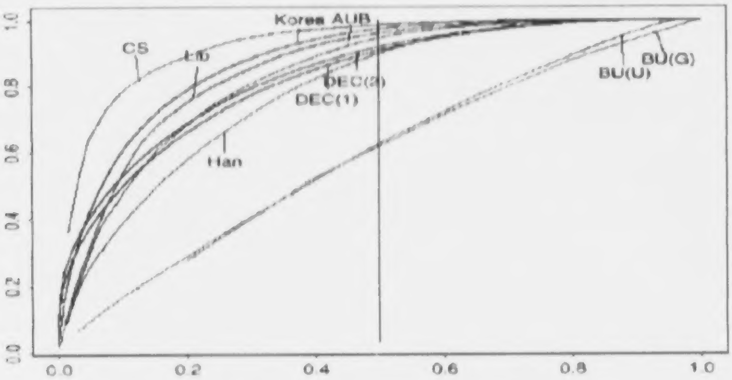


Figure 3: Clients concentration of accesses.

Fig. 4 shows that 90% of the bytes transferred come from 25% (see vertical line in figure) of the accessed servers. This is invariant 8 in Tab. 2.

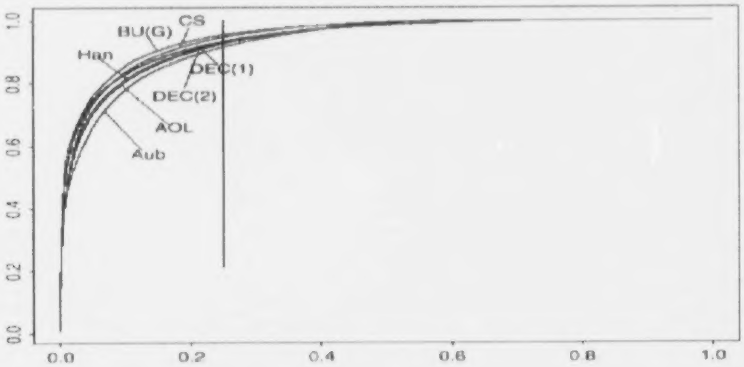


Figure 4: Server concentration of bytes accessed.

Success Rate

Retrieving documents successfully from Web servers happens with a high percentage in all workloads. We assume that a file is retrieved successfully if the Web server returns one of the following status codes in HTTP: 200 *success*, 304 *not modified* or 204 *OK but no contents*. In all workloads we notice this percentage is in the range 88%-99%. We include this invariant in Tab. 2 as invariant 9. The 304 return code is of particular interest for us since it reflects the percentage of files that are retrieved from the local cache. Interestingly the status code 400 *client error (bad request)* is zero in all workloads except the DEC workload, where the error appears 1.8%-2% of the time.

We also tested if the previously identified invariants hold for the VT-CS workload over a year period. To do this we split the workload into monthly log files and applied the same analysis that we used to identify the invariants across workloads. We used the VT-CS log file because it has the longest duration since it spans over 11 months. Although the conclusions drawn from such an analysis cannot be generalized since we are using one workload, still we can identify trends or changes that appear over time. Regarding all the invariants in Tab. 2 there is no change over time.

Summary and Conclusions

Although WWW users represent different user groups and different backgrounds, they have common behavior with respect to the invariants identified in Tab. 2. The identified shared behavior is important for two reasons, first, we can use it to generalize and come up with statistical models that can be used for simulation and modeling studies. For example, the identified distributions in Figures 1-4 and in the identified statistics for file types and sizes can be used to generate synthetic proxy workloads for simulation studies. Second, we can use some of the identified invariants to make conjectures about caching and prefetching. For example, the high locality of reference encountered in the examined workloads suggests that caching and maybe prefetching of documents are potential factors in solving the WWW scalability problem.

References

[Abdulla et al. 1997] Ghaleb Abdulla, Edward A. Fox, Marc Abrams and Stephen Williams (1997). WWW proxy traffic characterization with application to caching. *Technical Report TR-97-03*, Dept. of Comp. Sci., Virginia Tech, 660 McBryde Hall, Blacksburg, VA 24060.

[Abrams & Williams 1996] Marc Abrams and Stephen Williams (1996). Complementing surveying and demographics with automated network monitoring. *WWW Journal* 1, (3), 101-119, July 1996.

[Arlitt & Williamson 1996] M. F. Arlitt and C. L. Williamson (1996). Web server workload characterization: The search for invariants. *Proc. SIGMETRICS*, Philadelphia, PA, April 1996. ACM, 160-169.

[Crovella & Bestavros 1996] Mark E. Crovella and Azer Bestavros (1996). Self-similarity in World Wide Web traffic evidence and possible causes. *Proc. SIGMETRICS*, Philadelphia, PA, May 1996. ACM, 126-137.

[Cunha et al. 1995] Carlos R. Cunha, Azer Bestavros, and Mark E. Crovella (1995). Characteristics of WWW client-based traces. *Technical Report BT-CS-95-010*, Dept. of Comp. Sci., Boston Univ., 11 Cummington St, Boston, MA 02215, July 1995.

[Digital 1996] Digital Equipment Corporation (1996). Digital's Web proxy traces URL:ftp://ftp.digital.com/pub/DEC/traces/proxy/webtraces.html.

[Stevens 1996] W. Richard Stevens, (1996) *TCP/IP Illustrated*, Vol. 3, Reading, Massachusetts, Addison-Wesley.

[Wooster & Abrams 1997] Roland Wooster and Marc Abrams (1997). Proxy Caching That Estimate Page Load Delays. *Proc. 6th International World Wide Web Conference*, May 1997, 325-334.

Acknowledgments

We thank Dan Aronson from America Online, Digital Equipment Corporation, Seoul National University, and Boston University for making their log-files available. Members of the Virginia Tech NRG provided helpful comments on the manuscript. NSF grants NCR-9627922 and CDA-9312611 partially supported this work.

Business Use of the Internet in New Zealand: A Follow-Up Study

Walt Abell, Centre for Computing and Biometrics, Lincoln University, New Zealand, abell@lincoln.ac.nz

Steve Black, Information Tools Ltd, Auckland, New Zealand, sblack@xtra.co.nz

Abstract: A survey of New Zealand businesses using the Internet was undertaken in 1995 and a follow-up survey carried out in 1996. Both surveys looked at current and expected uses, perceived benefits and problems areas of Internet use by business. Interesting results include: a substantial increase in providing on-line customer services, small and technology focused companies making more use of the Internet; and an increase in problems with the technology and Internet Service Providers.

Introduction and Background

Many reasons and strategies for business use of the Internet have been proposed and discussed in the media and the IT industry (e.g., [Cronin, 1995] and [O'Reilly, 1996]). However, much of what has been written is anecdotal and only highlights successful cases. There are a growing number of studies (such as [Pitkow & Kehoe, 1994-1996]) that attempt to quantify individual consumers' use of the Internet by gender and age, purchasing preferences, etc. However, only a handful of studies have looked at how businesses are using the net (and why) and most of these have been carried out by market research companies who charge hefty fees for the information (e.g., [Peck, 1996]).

New Zealand is a small country which is geographically isolated from most world markets. The Interenet has the potential to enable New Zealand businesses to compete on a more even footing with their larger overseas competitors. In New Zealand, as in other countries, the adoption of the Internet by businesses has increased rapidly, as evidenced by the large increase in commercial domain name registrations [McDonald, 1997]. A whole support industry has sprung up to help businesses devise and implement their Internet plans.

In order to get a picture of how New Zealand businesses were using the Internet, a study was conducting in 1995 [Abell and Lim, 1996]. A survey approach was used to look at:

current and future usage of the Internet
reasons for and perceived benefits of Internet use
use of the Internet for marketing and advertising
problems and issues associated with Internet use

The current study involved recontacting the same companies in 1996 (15 months later) to see how their usage and perceptions had changed.

Methodology

The follow-up survey was kept as similar as possible to the original to enable comparisons to be made. However, some items were removed and a few new responses and questions added. The original and follow-up questionnaires are available (along with full result tables) at http://www.lincoln.ac.nz/ccb/staff/abell.htm .

An initial request for participation was sent by e-mail to the 116 respondents to the original survey. While this produced some immediate replies, the rest had to be contacted by phone (where this was possible). There were a number problems with the delivery and forwarding of messages (which also occurred in the original survey). This is an interesting result in itself and brings into question the usefulness of e-mail for surveys or indeed for making positive contact with companies. A summary of the responses to the participation requests is given in [Tab. 1].

Response	*N*	%
Agreed to participate	81	70
Declined to participate	3	3
Not contactable (details not provided in original survey)	19	16
Did not reply to e-mail or phone message	4	3
Ceased trading	7	6
Ceased using the Internet	2	2
Total	116	100

Table 1: Responses to participation requests

As in the original study, respondents were given the option of how to receive the questionnaire; well over half opted for e-mail. This is a good illustration of the change in Internet usage as only a small number in the original study opted for e-mail.

A total of 68 forms were returned. One was a duplicate and two could not be matched to an original questionnaire, leaving 65 valid responses (56% response rate).

Sample characteristics

As with the original survey, the sample was self-selected so the results obtained cannot be generalised to the wider business population. However, it is useful to look at the similarities between the follow-up and the original study groups. A breakdown by size and technology focus for both groups is given in [Tab. 2].

Original Group			
Company Size	***Non-Tech Focus****	***Tech Focus****	***Totals***
1-50	27%	47%	73%
>50	19%	7%	26%
Unknown	0%	1%	1%
Totals	46%	54%	100%

Follow-up Group			
Company Size	***Non-Tech Focus***	***Tech Focus***	***Totals***
1-50	28%	43%	71%
>50	20%	9%	29%
Unknown	0%	0%	0%
Totals	48%	52%	100%

*3 respondents changed their technology focus response from the original study

Table 2: Breakdown of original and follow-up groups

It appears that the follow-up group does not differ markedly from the original on these variables. Further comparisons of the original survey responses of both groups show little difference except in the area of marketing and advertising on the Internet (due to small numbers involved). Since the two groups are sufficiently similar, only the responses from the follow-up group to both surveys will be considered in the rest of this paper.

Survey Results

As could be expected, most current uses and benefits increased as shown in [Tab. 3] and [Tab. 4]. There was a corresponding drop in uses and benefits expected within the following twelve months.

Use	*Original*	*Follow-up*	*Change*
To get information from suppliers	65%	83%	18%
Provide information to customers	45%	69%	24%
Send orders to suppliers	37%	42%	5%
Receive orders from customers	34%	46%	12%
Market & product research	40%	58%	18%
E-mail Communications	91%	94%	3%
R&D/ Sharing of software, data or information	48%	55%	7%
Advertising job vacancies	11%	23%	12%
To be seen at the forefront of technology	54%	51%	-3%
Marketing and advertising	28%	55%	27%
Voice or video conferencing	2%	6%	4%

Table 3: Current uses

Benefit	*Original*	*Follow-up*	*Change*
Lower cost of obtaining supplies	20%	32%	12%
Faster, more flexible delivery from suppliers	31%	38%	7%
Better service and support from suppliers	51%	57%	6%
Increase in market share	22%	18%	-4%
Lower cost margins	23%	20%	-3%
Greater customer satisfaction	34%	55%	21%
Ability to reach international markets	38%	45%	7%
Effectiveness in information gathering	78%	80%	2%
Increased productivity	42%	46%	4%
Availability of expertise regardless of location	57%	57%	0%
Better awareness of the business environment	32%	42%	10%
*Improved communications**	9%	77%	

* response to "Other" in original survey, added as listed choice in follow-up survey

Table 4: Current benefits

Customers and Suppliers

Of the major uses, there was a large jump in both providing information to customers and receiving orders on-line. This was mirrored by a substantial increase in the "greater customer satisfaction" benefit. It is not clear whether the companies actually measured customer satisfaction or whether they assumed that providing better on-line access would lead to greater satisfaction. Interestingly, there was a drop in the "increased market share" benefit which could indicate that the companies were focusing on their existing customer base.

On the other hand, there were smaller increases for getting information from suppliers and ordering from suppliers. The latter was one of the few areas where expected use did not decline with 30% of respondents expecting to do this in the next year. There were also small increases in the "lower cost of obtaining supplies" and "better service and support from suppliers" benefits. The lower cost sentiment may be partly explained by the increase in market and product research, with the ability to compare prices and products more effectively. However, these results contrast with the small drop in the lower cost margins benefit. It would be interesting to study the companies claiming lower costs in more detail to see how the Internet actually impacts on this.

Marketing and Advertising

The number of companies using the Internet for marketing and advertising doubled in the follow-up survey. Almost all companies used a home page for this purpose and 56% used ads or links on other web sites. A greater number of companies in the follow-up survey kept statistics on customer visits to their site but the level of analysis varied widely from very detailed to "trying to make sense of them".

Impediments to Internet Use

While increased uses and benefits were expected, a rise in some of the problem areas of Internet use was a surprising result as shown in [Tab. 5]. However, there was a drop in the "suppliers and customers not connected" response which is consistent with the rapid growth of the Internet. A smaller drop in the "difficulty in locating information" response may be a result of improving search facilities and/or increasing user sophistication.

Reasons	*Original*	*Follow-up*	*Change*
Technical limitations of hardware/software	28%	49%	22%
Lack of expertise or personnel	23%	32%	9%
Suppliers/Customers not connected	72%	57%	-15%
Difficult to locate information	34%	28%	-6%
Connection and/or usage charges too high	20%	23%	3%
*Problems with ISP**	12%	22%	
*Concerns about security**	3%	42%	

* response to "Other" in original survey, added as listed choice in follow-up survey

Table 5: Reasons for not benefiting

The increases in the technical limitations and lack of expertise responses indicate that Internet technology still has some way to go in terms of useability. It could also be due to companies attempting more ambitious projects requiring more sophisticated skills (e.g., CGI or Java programming). The high response to security concerns indicates that there were still doubts in this area despite the advances made in encryption, etc. This may well be an education and public relations issue rather than just a technical one.

The problems with ISP response has some serious implications for the industry. The New Zealand IT media has chronicled a series of pricing and service problems with ISPs [Hosking, 1996]. Indeed, some of the problems in using e-mail to contact survey participants may have been caused by ISP problems.

Companies not using the Internet for marketing also indicated technical limitations and lack of expertise as reasons. In addition, there was a small increase in those who did not think that Internet marketing was effective (12% to 21%). Of the two new responses added to the follow-up survey, 45% selected not having time to research and set up a system while 24% said that their company had no policy on Internet use.

Internet issues of concern (security, frivolous use, etc) had similarly high ratings in both surveys. The only substantial change was a very high (98%) response for the system being reliable.

The rating for overall effectiveness of the Internet was very similar in both surveys. However, the specific rating for effectiveness for Internet for marketing and advertising was slightly lower in the follow-up group.

Company type and Internet use

A higher reporting of uses and benefits by small and/or technology focused companies was present in both the original and follow-up surveys. However, the gap between large and small companies was smaller in the follow-up while the difference between technology and non-technology companies was the same or greater. It is important to keep in mind that there was an overlap between the two groups. Companies that were both small and technology focused reported the most Internet uses and benefits.

While the technology result is not surprising, the small/large difference is at odds with overseas trends. O'Reilly and Associates continue to report that Internet uptake by large North American companies far exceeds that by smaller ones [Peck, 1996]. The New Zealand situation is of course quite different, with almost all businesses considered "small" (less than 100 employees). There is also a perception that New Zealanders are quick to adopt new technology. The small/large gap possibly reflects a more flexible attitude to experimentation by smaller businesses. The narrowing of that gap may mean that larger companies have become aware of the potential benefits (and the growing imperative) to be on-line.

Summary

Although this study used a self-selected sample, it does point to some interesting trends in Internet use by New Zealand businesses including an emphasis on customer service (which overshadows marketing) and a steady move toward on-line transactions. From a New Zealand point of view, the Internet provides businesses with an unparalleled opportunity to reach distant markets. However, the continuing concerns over security and technical (and ISP) problems could hamper companies' plans. Further research in these areas is crucial if New Zealand is to make the most of the full potential of the Internet for electronic commerce.

References

[Abell and Lim 1996] Abell, W., & Lim, W. (1996). Business Use of the Internet in New Zealand: An Exploratory Study, 1996, *Second Australian World Wide Web Conference*, 1996, Southern Cross University Press, Lismore, NSW, Australia, 33-39.

[Cronin, 1995] Cronin, M. (1995). *Doing More Business on the Internet.* New York: Van Nostrand Reinhold.

[Ellsworth and Ellsworth, 1995] Ellsworth, J., & Ellsworth, M. (1995). *Marketing on the Internet: Multimedia Strategies for the World Wide Web.* New York: John Wiley & Sons Inc.

[Hosking, 1996] Hosking, R. (1996). Customer forgotten in ISP row. *New Zealand Computerworld*, 474, 1-2.

[McDonald, 1997] McDonald, M. (1997). Corporate Internet explosion. *New Zealand InfoTechWeekly*, 284, 1-2.

[O'Reilly, 1996] O'Reilly, T. (1996). Publishing Models for Internet Commerce. *Communications of the ACM*, 39 (6), 79-86.

[Peck, 1996] Peck, R. (1996). *Conducting Business on the Internet*. O'Reilly and Associates, Inc. http://www.ora.com/www/research/netcraft/

[Pitkow and Kehoe, 1994-1996] Pitkow, J., & Kehoe, C. (1994-1996). *Georgia Institute of Technology WWW User Surveys*. http://www.cc.gatech.edu/gvu/user_surveys/

Easy Ed: An Integration of Technologies for Multimedia Education

G. Ahanger and T.D.C. Little
Multimedia Communications Laboratory
Department of Electrical and Computer Engineering
Boston University, Boston, Massachusetts 02215, USA
(617) 353-8042, (617) 353-6440 fax
{*gulrukh,tdcl*}*@bu.edu*

Abstract: The accessibility of the World Wide Web and its flexibility for conveying digital information in various forms makes it a convenient mode of communication for education. In this paper with the help of a distance learning application called "Easy Ed," we demonstrate how these properties of the World Wide Web along with a *data model* can be used to provide a classroom environment on the Internet. Easy Ed provides a rich medium for education that is achieved by integrating information across the different media types (text, video, audio, and graphics) in hyper-media form. Metadata conforming to the data model about different media types is stored in a relational database, which not only facilitates authoring but also makes it possible to reuse existing instructional material. Another unique concept of Easy Ed is the *dynamic repurposing* of content at the time of access. Dynamic information generation helps to customize information according to a user's level of comprehension, the information medium, and hardware compatibility.

Introduction

Distance education involves providing a user with instructional material for self or group learning for geographically dispersed students. The basic outline for instructional material can be established by domain experts that will remotely supervise a student. Distance education is not meant to replace the instructor or other experts but to let a larger audience benefit from their expertise. From the student's perspective it is much more convenient if the user can view information at convenience both with respect to time and duration of viewing. As an outcome of a study conducted by Wetzel et al. [Wetzel et al. 1994] on the effectiveness of video as a learning medium it was noted that the medium of education should be non-linear and dynamically paced. For example, a student can follow links for additional information on a particular topic and then continue with better understanding. As the information can be accessed remotely, the World Wide Web (WWW) and the Internet make information time- and place-independent [Harasim 1990]. Access to geographically isolated communities, multiple participation, and sharing of diversity and similarity among people can also be added to the benefits of distance education via the Internet [Schrum and Lamb 1996].

A number of experiments [Hiltz 1995, Rada 1996, Schrum and Lamb 1996] have been conducted to assess the effectiveness of distance education as compared to a conventional classroom environments. The results show that the mastery of material of students using digital libraries is equal or superior to that of a traditional classroom. Students are able to better synthesize or establish relationships between diverse ideas. These results are judged successful especially with students who worked full time or those who were geographically scattered. Though such experiments are a success they do not utilize the full capabilities of the WWW and the Internet, i.e., the capability of providing true multimedia information. To further benefit from multimedia technology, we need to integrate information of diverse forms including video, audio, text, and images to provide a richer hyper-linked medium for learning. For example, in addition to textual information about how to perform a chemistry experiment, one might provide a link to a video clip of an expert demonstrating the experiment.

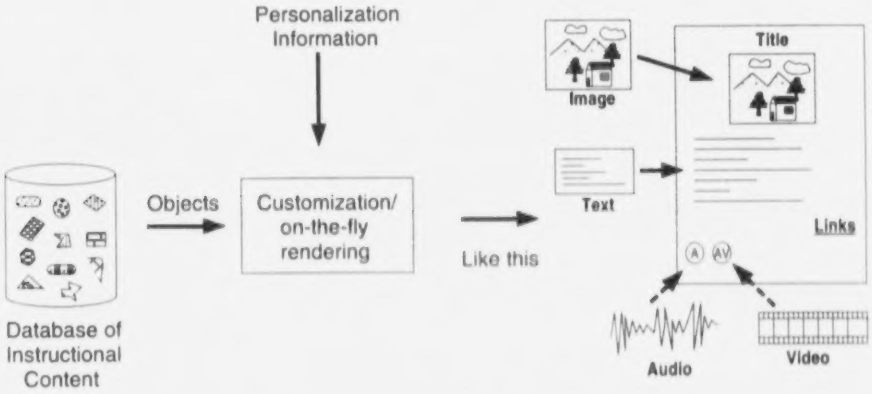

Personalization Information
Objects
Customization/ on-the-fly rendering
Like this
Database of Instructional Content
Image
Text
Title
Links
A
AV
Audio
Video

Figure 1: Customization and Page Composition

Easy Ed, a distance learning application, is a result of our investigation of technologies for integrating educational material from various media types. In addition to being a true multimedia distance learning application, Easy Ed has a number of novelties. First, it dynamically generates a course from metadata stored in a relational database on-the-fly. Various related multimedia objects are integrated at the time of rendering information, i.e., the information is not pre-composed [Fig. 1]. This not only makes it easier to reuse the objects but at the same time reduces the opportunity for material to be reproduced en masse. Second, this technique eliminates the need for data replication (e.g., if the same instance of text is to be displayed in two different topics we only require a single instance in our archive, whereas pre-composed static documents of the same text requires replication). Not replicating the data makes a considerable difference in storage savings for large instructional content (e.g., video). Third, the use of ***dynamic document generation*** helps in customization of information. Depending on a user's preferences the information can be easily filtered to reduce excess content (e.g., if a network capacity does not allow realtime delivery of video then this medium can be omitted). Fourth, authoring is simplified as an author can form a new course from existing information by identifying relationships between different objects. Fifth, an effective medium of learning is provided in Easy Ed by integrating concepts of different media types. Finally, we have simulated the look and feel of a conventional book but with the incorporation of content-based *tours* and searching.

In addition to providing an integrated environment for education, we are also motivated by our desire to reuse legacy video-based instructional materials existing in our own lab. Some of the specific objectives in this effort are:

- Offer dynamic or self-paced education.
- Provide non-sequential access for improved learning.
- Provide tailored material for individual needs.
- Save costs of creation and delivery.
- Allow courses not offered in a semester to be made available.
- Allow access to related courses.
- Allow remote access.

Hence, with these objectives and a desire for a true multimedia application, we set out to create the distance learning application called Easy Ed.

BEST COPY AVAILABLE

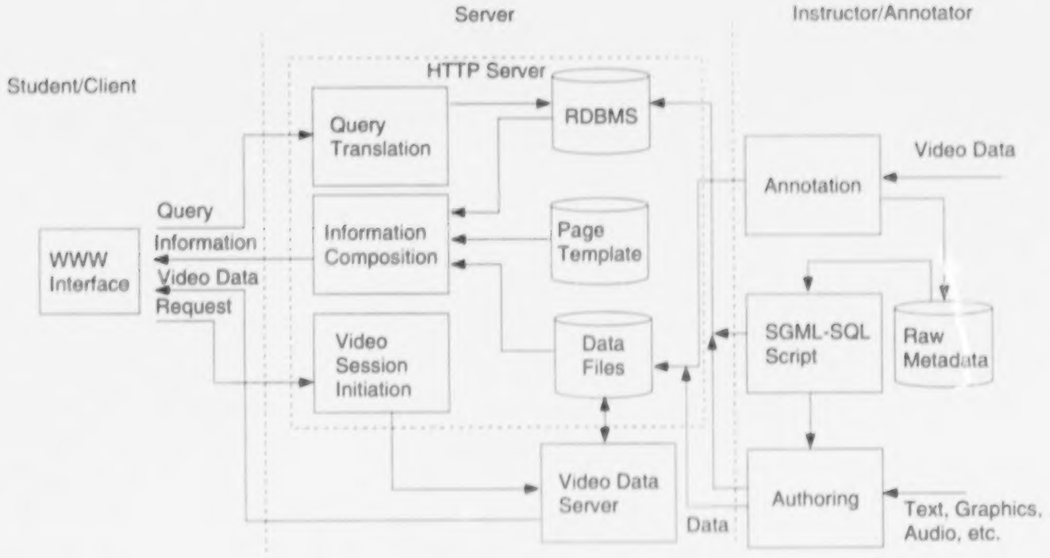

Server
Instructor/Annotator
Student/Client
HTTP Server
Query Translation
RDBMS
Video Data
Annotation
Query
Information
Information Composition
Page Template
WWW Interface
Video Data
Request
Video Session Initiation
Data Files
SGML-SQL Script
Raw Metadata
Video Data Server
Authoring
Data
Text, Graphics, Audio, etc.

Figure 2: System Architecture of Easy Ed

Architecture and Features of Easy Ed

The architecture of Easy Ed can be divided into three parts; instructor/annotator, student/client, and server. The annotator extracts information from raw data based on an instructional data model. The extracted information is then stored in a relational database. The client component provides a student with means of access to the stored information. The server deals with processing client requests, searching, and composition of data prior to delivery to the client.

The unique features (e.g., dynamic document generation) supported by Easy Ed are a result of the data model and composition process. The instructional data model is based on *context*. A random segment of a topic is not enough to comprehend the meaning of what is being said completely, a context has to be established. Therefore, the unit of information rendered is in the form of a *topic* and a course is offered as a set of topics. Each topic can be composed of graphics, text, and hyper-links to information in the form of text, video, graphics, or audio. Providing links to related video segments achieves non-linear and self-paced viewing of video. The informational components are treated as objects, a single object can belong to multiple topics, and a topic (single or multiple instance) can belong to multiple courses. For example, consider a topic being taught in two courses or in the same course but at different times. Therefore, we achieve different instances of the same topic. We consider each instance of the topic as a separate identity but with conceptual association. Some of the important capabilities supported by the data model are as follows:

- **Customization:** The database is designed to limit access after authoring by only allowing delivery of a subset of the objects to the client on request. This is easily achieved by treating the contents of the instructional database as distinct objects and combining these objects at the time of rendering. In the instruction database the objects are "page," "graphics," "video," "transcript," "audio," and "links" [Fig. 3]. Each topic is composed of pages which can be viewed sequentially to establish context and provide controlled information. The page object is a container for objects it contains (i.e., graphics, text, references, audio, and video).
- **Tours:** The ordering of the presentation of the topics in a course can be changed by changing the order in the relational database, thereby generating different "tours" for the same course. [Fig. 4]

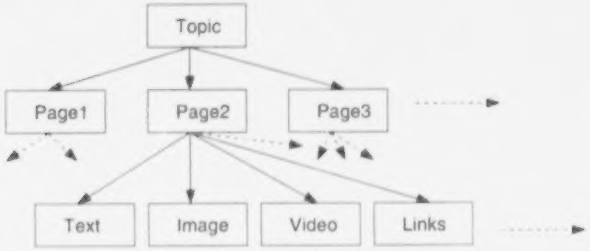


Figure 3: Object Hierarchy in a Topic

depicts a scenario of tour formulation for a particular topic. Tours are useful if a course can be offered at varying levels of difficulty (e.g., beginner, intermediate, and advanced).

- **Fast Access:** Components such as abstract, transcript, related text, audio, or video systems streams are used to provide information at different granularities. A user can browse through the database using the concepts provided or using a complete keyword search, accessing at any of these granularities.
- **Authoring & Repurposing:** The authoring of existing courses or any new course is simplified by the data model. An instructor can identify new relationships between objects to create a news topic or a course. An instructor does not have to manually assemble information. Not only existing informational material (e.g., images, text, video clips) can be used for composing new courses but any new material can be easily added as objects to the database for integration into a course.

Operation

On initial access, a student can browse the database by "Course," "Topic," "Instructor," and "Year." If the search is made by course name/number then the system lists titles and creation dates of all courses in the database satisfying the query. When a student chooses a course then the system generates a *view* of that course and displays it to a student as shown in [Fig. 5]. A view displays the course and various available tours (e.g., beginner, intermediate, and advanced) associated with it. Once a student selects a view, all the topics offered in the view are displayed and by clicking on a particular topic the contents are displayed. The browse mechanisms for "Topic," "Year" and "Instructor" operate in a similar manner.

In addition to browsing the database, a student can search for particular content in the database. The student can search using a form-based interface with details about the "Course ID," "Course Title," "Year," "Topic," "Instructor," and "Session." The student can fill in any one of the fields or any combination of these fields. To provide a more detailed search, a search based on "keywords" can also be executed.

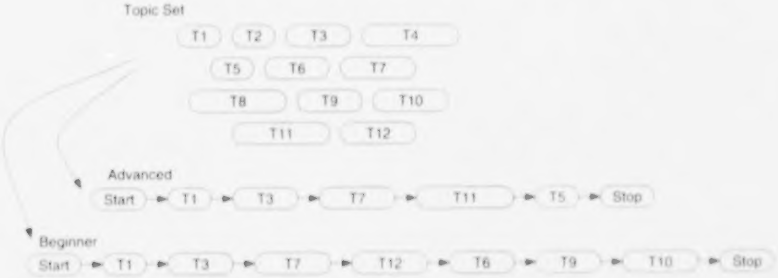


Figure 4: An Example of Tour Formation

Query Processing

Once a student issues a query (e.g., clicks on a certain topic in the Course View Interface), a Common Gateway Interface (CGI) script of the WWW server is executed translating the query into the Structured Query Language (SQL) format and sends the query to a relational database as shown in [Fig. 2]. The system finds all page objects contained in the selected topic and then finds the objects contained within each page. The retrieved information is sent through an information composition module which composes the information according to the template provided. If a topic has more than one page then a series of pages are composed and are dynamically linked together and delivered to the student. Graphics and text are rendered in a WWW interface and links are provided to any relevant video or audio clips.

Implementation

We use a relational database called Mini SQL (mSQL) [Hughes 1995] as a database interface from a RDBMS to HTTP server. The database interfaces with the WWW by the C language API of mSQL. Video indexing is performed using a graphical annotation tool called Vane [Carrer et al. 1997]. The metadata are stored in conformance with the SGML format tailored to video data as specified for Vane. The database is automatically populated with metadata from the SGML files with the help of scripts written in Perl 5.

The client is written using HTML and JavaScript. Because the URL addresses are resolved on-the-fly, utilizing JavaScript is very convenient. A WWW browser is used to display the images, text, and audio. To play video, a student initiates a streaming session by a click on a video icon. Streaming is implemented using our own protocol which achieves a small start up latency and lossless delivery. The video is displayed in a separate window.

Summary

This research is based on our investigation of technologies for digital video archival and distribution. We have created a hyper-media environment for distance learning by linking small, cohesive units of video data with text. This not only provides important visual information but at the same time allows self-paced education.

The data model is simple and flexible because coherent information units are treated as objects. Dynamic assembly of information at the time of rendering makes the process of customization straightforward. Objects are incorporated or deleted depending on a student's preferences or the network's and client's capabilities, thereby providing fast access to information at various granularities. Dynamic repurposing not only allows an object to be part of different courses simultaneously but achieves storage savings; objects are replicated only at the time of rendering. Students with various levels of expertise can be serviced by different tours of a course by storing different sequences of topics in the relational database.

Thus, Easy Ed in addition to having a look and feel of a conventional book, efficiently integrates information in multiple media. It provides a flexible access to information while accommodating student preferences in a platform-independent manner.

Acknowledgment

This work is supported in part by the National Science Foundation under Grant No. IRI-9502702.

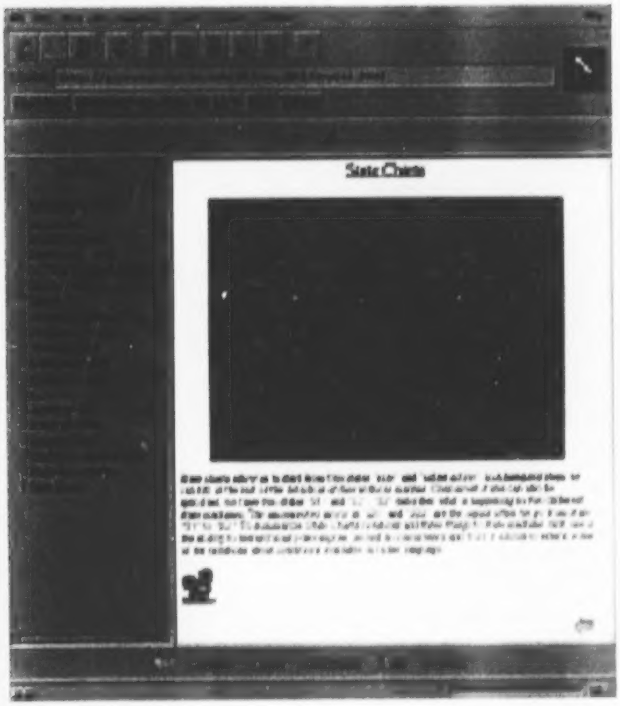


Figure 5: Interface for Display of Course Views and Contents

References

[Carrer et al. 1997] Carrer, M., Ligresti, L., Ahanger, G., & Little, T.D.C. (1997). An Annotation Engine for Supporting Video Database. *Journal of Multimedia Tools and Applications*, 5 (3).

[Harasim 1990] Harasim, L.M. (1990). Online Education: An Environment for Collaboration and Intellectual Amplification. *Online Education: Perspective on a New Environment*, New York: Pergamon. 39-64

[Hiltz 1995] Hiltz, S.R. (1995). Impact of College-Level Courses via Asynchronous Learning Networks: Focus on Students. *Sloan Conf. on Asynchronous Learning Networks*, 1995, Philadelphia.

[Hughes 1995] Hughes, D.J. (1995). Mini SQL: A Lightweight Database Engine. *QUESTnet*, 1995.

[Rada 1996] Rada, R. (1996). Teaching on the WWW: Assignment Focus and Information Indexing. *Journal of Universal Computer Science*, 2 (10), 732-743.

[Schrum and Lamb 1996] Schrum, L., & Lamb, T.A. (1996). Groupware for Collaborative Learning: A Research Perspective on Processes, Opportunities, and Obstacles. *Journal of Universal Computer Science*, 2 (10), 717-731.

[Wetzel et al. 1994] Wetzel, C.D., Radtke, P.H., & Stern, H.W. (1994). *Instructional Effectiveness of Video Media*. Hillsdale, NJ: Lawrence Erlbaum Associates.

BEST COPY AVAILABLE

A User-Adapted Interface for a Search Engine on the World Wide Web

Leonardo Ambrosini, Vincenzo Cirillo and Alessandro Micarelli

Dipartimento di Informatica e Automazione
Università di Roma Tre
Via della Vasca Navale 84
00146 Roma, Italy
{ambrosin, cirillo, micarel}@inf.uniroma3.it

Abstract

In this paper we describe a system for Information Filtering on the World Wide Web based on User Modeling. The system is capable of selecting HTML/Text documents, collected from the Web, according to the interests and characteristics of the user. We have used the system as an intelligent interface for the search engine AltaVista™. The system is composed of two main modules: (1) a User Modeling module, based on a hybrid architecture, capable of building a representation of the user's interests and characteristics (User Model), and (2) an Information Filtering module that takes advantage of a semantic network and well-structured database information. The system has been implemented in Java™ on a Pentium™-based platform.

1. Introduction

The growth of Internet and the World Wide Web (WWW) makes it necessary for the end-user to cope with huge amounts of information readily available on the net. There are at present many search engines on the market, suitable for extracting information across the net. One of the problems posed by the use of such tools is that very often the information retrieved proves to be too vast and too generic to be immediately useful to the user, who must sift through it manually -- a tedious job at best. This is especially true when the keywords used for the query are common terms. As a consequence, filtering information (Belkin & Croft, 1992) on the Web is an increasingly relevant problem. In this paper we present a system for Information Filtering of HTML/Text documents collected from the WWW, where the selection of the documents relevant for a particular user is performed on the basis of a model representing the user's interests and characteristics. As a first application, the system has been used as an intelligent interface for AltaVista™ (both on *advanced* and *simple* query modalities), the well-known search engine designed by Digital Inc. to extract any kind of information across Internet.

The paper is organized as follows. In the next Section the general architecture of the system is presented. Section 3 describes the User Modeling component, based on a hybrid architecture, which identifies user characteristics. Section 4 presents the Information Filtering component, describing the algorithms used to filter the relevant documents, according to the model of a particular user. In Section 5 we conclude by presenting some experimental results obtained from using our system in real-life situations.

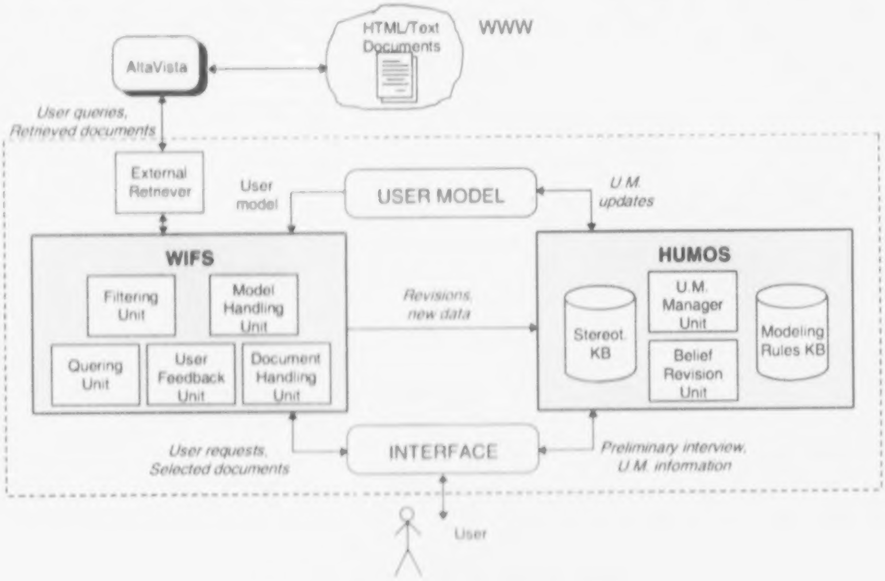

HTML/Text Documents
WWW
AltaVista
User queries, Retrieved documents
External Retriever
User model
USER MODEL
U.M. updates
WIFS
Filtering Unit
Model Handling Unit
Quering Unit
User Feedback Unit
Document Handling Unit
Revisions, new data
HUMOS
U.M. Manager Unit
Stereot. KB
Belief Revision Unit
Modeling Rules KB
User requests, Selected documents
INTERFACE
Preliminary interview, U.M. information
User

Figure 1: General architecture of the system.

2. General Architecture of the System

Figure 1, which is self explanatory, shows the two main components of the system: the Information Filtering Subsystem (WIFS, Web-oriented Information Filtering System) and the User Modeling Subsystem (HUMOS, Hybrid User Modeling System). For each component, we describe the Knowledge Bases (KB) and main units providing the basic functionality of the system. Both subsystems have been purposely designed as domain independent, i.e. they are intended as shell systems that can be used on domains different from those they were originally developed for (To give an example, the HUMOS subsystem is currently being integrated into a case-based training system: see Papagni, Cirillo & Micarelli, 1997). This is the reason why units strongly depending on the document domain to be searched have been kept outside the design of the two subsystems: in particular, the user interface and the unit charged with searching for information on the net (the External Retriever).

The system has been implemented in Java™ on a Pentium™-based platform and it is composed of 15,000 lines of Java code, 1 MBytes of byte-code and 150 Java classes (external libraries not included).

3. The User Modeling Subsystem HUMOS

The User Modeling Subsystem uses an approach for user modeling based on *stereotypes* (Rich, 1983). A stereotype can be viewed as a description of a prototypical user of the class represented by the stereotype.

We have used a Method for the integration of symbolic Artificial Intelligence (AI) and artificial neural networks for the task of automatically inferring user stereotypes during the user modeling phase. In particular, our approach integrates an artificial neural network in a case-based reasoner. Case-based reasoning (Kolodner, 1993) is an analogical reasoning method. It means solving new problems by using old experience embedded in a data base of "cases". When a new problem to be solved (the new case) is input to the system there is an *indexing* phase, consisting in the retrieval of old cases that closely match with the new case, and an *adapt* phase, consisting in the adaptation of the old solutions to the new problem.

A possible case-based approach to the selection of the most suitable stereotype, on the basis of user behaviour, is presented in Figure 2. The case library contains the old cases (gathered from experts in the domain) in the form of frames, the slots of which constitute the "user description" (a pattern made up of the actual values of the attributes for a particular user), the "active stereotype" (that can be viewed as a pointer to the Stereotypes Library) and a *demon*, i.e. a procedural attachment, activated when the old case is indexed and which triggers the knowledge base of adaptation rules which fit the selected stereotype to the content of the user model. When the system is presented with a pattern of attributes relative to a particular user, the indexing module tries to find an old case that closely matches the new one (according to a given metric). The selected old case contains all the relevant information useful for classifying the user, i.e. the most suitable stereotype and the demon to be used to activate the adaptation rules, starting from the selected stereotype and the actual pattern representing user behaviour.

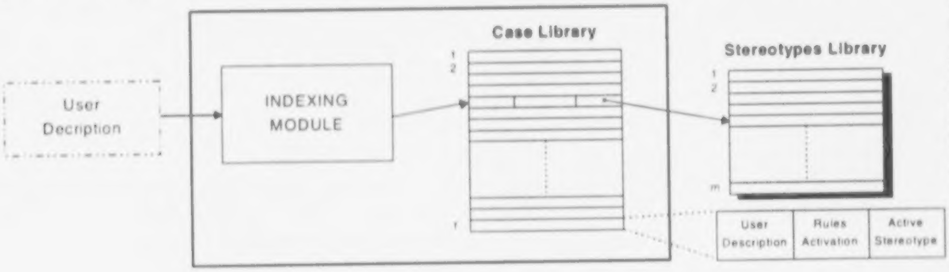


Figure 2: Case-based Approach to User Modeling.

One problem raised by this approach is in the determination of a metric to be used in the indexing module: in fact we have noticed that this type of classification of users must be made in the light of incomplete and often conflicting information. Our proposed solution (similar to a framework which has already been successfully experimented in the field of *adaptive hypermedia*; see Micarelli & Sciarrone, 1996a-b) consists of the use of a *function-replacing hybrid* (Fu, 1994; Goonatilake & Khebbal, 1995), where an artificial neural network implements (i.e., is functionally equivalent to) the module represented in bold line in Figure 2. The procedural attachment is not activated by the network but rather by the stereotype selected and the actual pattern determined. The old cases present in the case library are used as *training records* for training the network. As a result, the metric of the indexing module of Figure 2 is replaced by the *generalization* capability of the network. One advantage of this choice is that the distributed representation and reasoning of the neural network allows the system to deal with incomplete and inconsistent data and also allows the system to "gracefully degrade".

Since this kind of classification problem is, in general, not *linearly separable*, we have used a *Multi-Layer-Perceptron* (Rumelhart & McClelland, 1986) with three distinct layers. The first layer, the *input layer*, is composed of the neurons relative to the *n* attributes (that are coded into numeric values) present in all stereotypes. The *output layer* is composed of as many neurons as the number of the stereotypes. The output values are computed by the network according to a given input; this corresponds to the computation of a rank-ordered list of stereotypes present in the library. As for the *hidden layer*, there are no theoretical guidelines for determining the number of *hidden nodes*. We have selected the optimal number of hidden neurons in the context of the training procedure, where a *backpropagation* algorithm (Rumelhart & McClelland, 1986) has been used. During the training phase, we have used the *Simulated Annealing* algorithm (Kirkpatrick, Gelatt & Vecchi, 1983) for avoiding local minima as well as escaping from them when necessary (see Micarelli & Sciarrone, 1996a-b for more details concerning the authoring phase of the network).

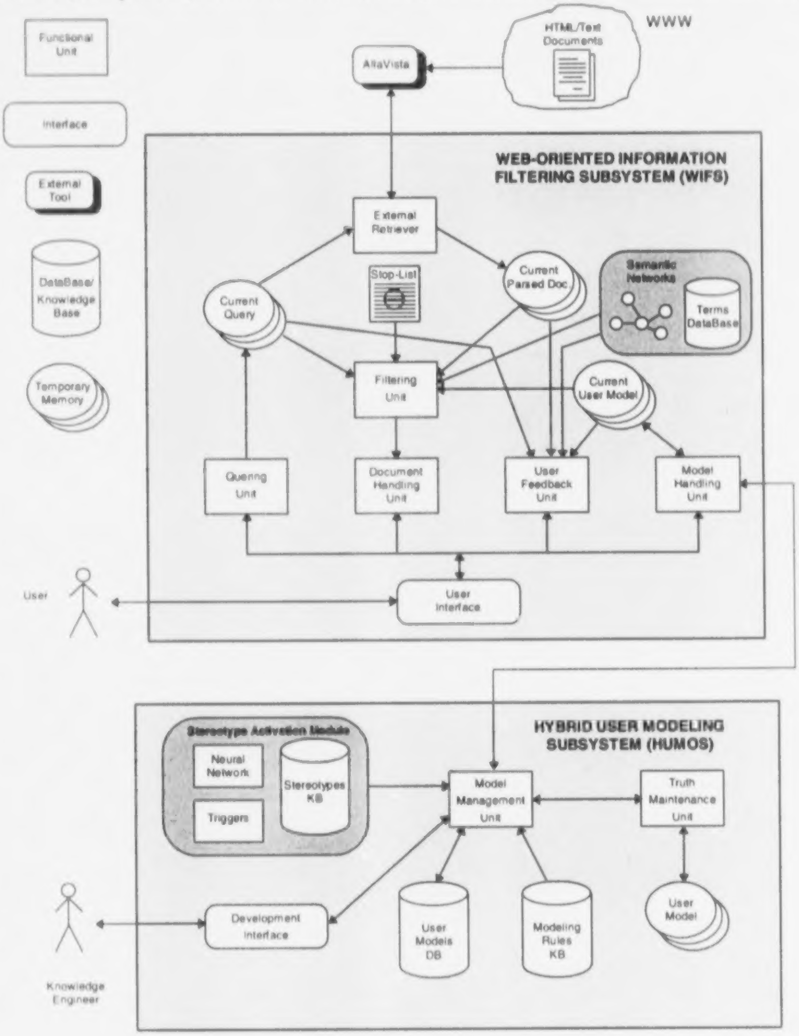

Functional Unit
Interface
External Tool
DataBase/ Knowledge Base
Temporary Memory
User
HTML/Text Documents
WWW
AltaVista
WEB-ORIENTED INFORMATION FILTERING SUBSYSTEM (WIFS)
External Retriever
Stop-List
Current Query
Current Parsed Doc.
Semantic Networks
Terms DataBase
Filtering Unit
Current User Model
Querying Unit
Document Handling Unit
User Feedback Unit
Model Handling Unit
User Interface
HYBRID USER MODELING SUBSYSTEM (HUMOS)
Stereotype Activation Module
Neural Network
Triggers
Stereotypes KB
Model Management Unit
Truth Maintenance Unit
Development Interface
User Models DB
Modeling Rules KB
User Model
Knowledge Engineer

Figure 3: Functional architecture of WIFS and HUMOS.

Bottom part of Figure 3 illustrates a simplified architecture of HUMOS, where main functional units and knowledge bases are shown to better understand how the overall modeling process works. The user modeling process entails the following activities: a) identifying current user; b) retrieving the proper user model, if any, or performing a preliminary interview; c) updating the model in order to insert or remove information about the user; d) retrieving data from the model.

Identification is actually performed chiefly by the host system. It gets login information from the user and hands it over to the *Model Management Unit* (MMU) that checks out whether the corresponding user model is available in the *DataBase of User Models*. If it is, the model is retrieved; otherwise a preliminary interview is conducted in order to get basic information about the user. Retrieving data on the basis of the model becomes almost a trivial operation whereas the operation of modifying the model becomes the crucial step. Before describing how the user model is updated, we should describe the data structure of the model itself a little better.

The User Model is made up of records containing information about the user. These are called *components* and look like t-uples with the following fields: *attribute*, *value*, *weight*, *semantic links*, *causal links*, *class*. Stereotypes are very similar to the User Model; in fact a stereotype is a list of components, but do not require any semantic or causal link or a class flag. A stereotype is nothing but a frame: *attributes* are the *slots* of the frame and for each slot the *value* facet corresponds to a list of value-weight pairs.

Whenever MMU is informed about new data to insert into the model, it tries to instantiate the data into the model. This is a critical operation which involves several tasks, entering a loop that ends when the model is considered stable and consistent. The loop basically involves the following activities: adding and removing

BEST COPY AVAILABLE

components from the model, checking the list of current active and disabled stereotypes, checking firing rules, looking for inconsistencies, resolving them and, finally, relooping back. The user model is considered stable and consistent when (1) the lists of current active and disabled stereotypes is not changed by the *Stereotype Activation Module* (SAM), (2) there are no new firing rules, (3) there are no inconsistencies. At this point it may be passed on to the host system.

The insertion of new components and the removal of old ones entails consistency checks. This is because new data may contradict old data. If an inconsistency is found, the MMU will then notify the *Truth Maintenance Unit* (TMU) (Doyle, 1979; Forbus & De Kleer, 1993) of the contradictory components. In addition the causal links have to be updated.

Inferential activity is carried out both by SAM (already described) and the *Inference Engine* (IE) that looks for rules the left-hand-side of which matches a component already present in the model. If any firing rule is found, its right-hand-side is adapted according to the match on the left-hand-side and is passed to the MMU to be inserted into the model.

Should any inconsistency be found, it is signalled to the TMU. TMU identifies components that justify the contradictory components and components justified only by the contradictory components, i.e., the *no-goods* (see also Brajnik & Tasso, 1994). According to rules encoded into the system, the inconsistency is then solved by selecting the components to be removed (never more than one). The no-goods depending on those components also have to be removed. During this activity TMU needs a *Working Memory*. The system interfaces have not been designed with the end user in mind, but rather to help a knowledge engineer to tune the system properly: user model editing capability is provided to expert users by the WIFS.

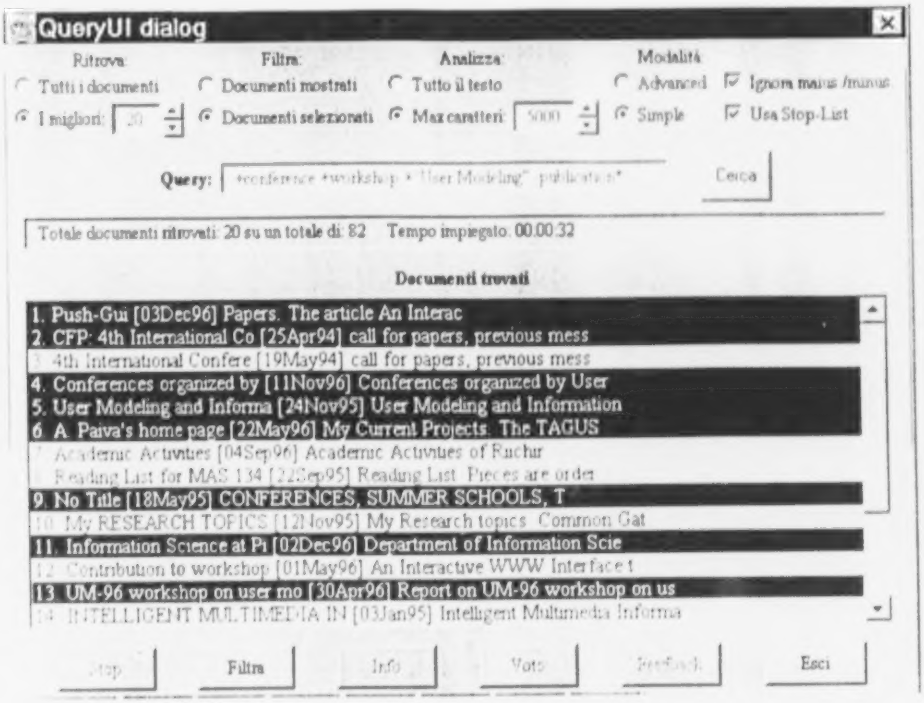
4. The Information Filtering Subsystem WIFS

As shown by the schema of WIFS architecture (Figure 3), the filtering process is made up of the following steps:

1. The Logon phase: identification of the Current User.
2. The Initialisation phase, performed by the Model Handling Unit: retrieval of the corresponding user model from HUMOS. Whenever the user is not known, the Unit conducts a preliminary interview in order to collect a first set of information. In particular it asks the user her/his likes and dislikes, assigning each an importance weight (positive for interesting attributes, otherwise negative). This information is then sent to HUMOS to identify a suitable stereotype and create an appropriate model.
3. The Editing/customising phase, performed by Model Handling Unit: user perusal and editing of the Model created for her/him. The User Model and the modifications performed by the user are sent to HUMOS in order to carry out the modeling process through the following activities: activation of stereotypes and firing rules and inconsistency checking. Moreover, the user can define multiple sub-models (called *classes*), one for each area of interest and may use each class to initiate a search. The sub-profiling is very important when queries are specific or complex, since some aspects of the generic user model may disrupt the filtering. Each model p is therefore a set of vectors p_C, one for each class C, as follows

   $$p_C = (\langle a_{C1}, v_{C1}, w_{C1}, s_{C1}, \ldots \rangle, \ldots, \langle a_{Cr}, v_{Cr}, w_{Cr}, s_{Cr}, \ldots \rangle)$$

   where $<a_{Ci},\ldots>$ are the components (see preceding Section), a_{Ci} is an attribute (of the domain known to the system), v_{Ci} its value (instance of a_{Ci}), w_{Ci} its weight (relevance factor of v_{Ci}), s_{Ci} are semantic links (components semantically related to v_{Ci}).
4. The Querying phase, performed by Quering Unit: display of a window where the user can input her/his query and establish the searching and filtering modalities. AltaVista syntax lets the user write both boolean queries (boolean AND/OR combinations of keywords) and structured queries (which allow the user to constrain matches to certain attributes, such as document type, title, host, etc.). Figure 4 is the snapshot of a typical query using the system: the user is looking for documents about "conference" and "workshop" on "User Modeling" (excluding "publications"). As one may see, at the top of the window the user has checked the parameters of interest to her/him: maximum number of documents to retrieve and to filter, AltaVista query modality, etc. The main text area contains the ranked list of documents retrieved by AltaVista (title, date and first line). The user can choose to filter all or a subset of these documents (in the example shown, a subset has been filtered).
5. The Parsing phase, performed by External Retriever: parsing, analysis and extraction of the structured representations of the HTML/Text documents retrieved by AltaVista. The structural elements are: title, abstract, author/s, URL, size, relevant keywords (with the help of the *Stop-List*) and their frequency in the text.
6. The Filtering phase, performed by the Filtering Unit: activation of the filtering algorithm we have proposed (*MAF*, Matching Algorithm for Filtering) in order to assign a *Score* to each representation calculated as the similarity between the document, the user model, and the query. It is interesting to note how MAF computes this



Ritrova: ○ Tutti i documenti ● I migliori: 20

Filtra: ○ Documenti mostrati ● Documenti selezionati

Analizza: ○ Tutto il testo ● Max caratteri: 5000

Modalità: ○ Advanced ● Simple ☑ Ignora maius./minus. ☑ Usa Stop-List

Query: +conference +workshop + "User Modeling" publication*

Totale documenti ritrovati: 20 su un totale di: 82 Tempo impiegato: 00.00.32

Documenti trovati

1. Push-Gui [03Dec96] Papers. The article An Interac
2. CFP: 4th International Co [25Apr94] call for papers, previous mess
3. 4th International Confere [19May94] call for papers, previous mess
4. Conferences organized by [11Nov96] Conferences organized by User
5. User Modeling and Informa [24Nov95] User Modeling and Information
6. A. Paiva's home page [22May96] My Current Projects. The TAGUS
7. Academic Activities [04Sep96] Academic Activities of Ruchir
8. Reading List for MAS 134 [22Sep95] Reading List. Pieces are order
9. No Title [18May95] CONFERENCES, SUMMER SCHOOLS, T
10. My RESEARCH TOPICS [12Nov95] My Research topics. Common Gat
11. Information Science at Pi [02Dec96] Department of Information Scie
12. Contribution to workshop [01May96] An Interactive WWW Interface t
13. UM-96 workshop on user mo [30Apr96] Report on UM-96 workshop on us
14. INTELLIGENT MULTIMEDIA IN [03Jan95] Intelligent Multimedia Informa

Figure 4: An interface of the system.

similarity: besides evaluating the conventional vector product (between corresponding vectors of document, profile and query), MAF properly exploits the occurrence of semantic links and terms (see below) found in the document. Supporting these structures provides a more accurate filtering process. All documents are ordered by descending *Score* and shown to the user by the Document Handling Unit. An important feature of MAF is the ability to identify topics composed of multiple keywords: in fact the system can distinguish between topics like "User Modeling" and single keywords like "User" and "Modeling".

7. The Feedback phase, performed by User Feedback Unit (UFM): input of a relevance value, assigned by the user, for each document viewed. The value tells the system how satisfied the user is with it. This feedback, the representation of the document, and the query, are then all used to modify the user model by inserting newly found topics and updating the weight of topics already in the model. Thus the model evolves according to user behaviour. The UFM also checks the weight of each component, and will delete any component whose weight is below a certain value; all attributes that are not "refreshed" by the user will be set aside. The updated user model is then sent to HUMOS in order to carry out the modeling process activities.

Let's take a look at the insertion of new components into the model, done by the UFM according to our proposed algorithm *SAF* (Semantic net/DB-based Algorithm for Feedback). If the system finds an unknown keyword *k* in a document, it first uses the *Terms DataBase* (TDB) to find the semantic meaning of the keyword. The TDB is structured to ease this task and evolves dynamically. If *k* is already in the TDB, then the model is updated by inserting *k*'s components, already known by the system: this dynamically broadens the semantic aspect of the model and its inferential capabilities as well. If *k* does not have a value in the TDB, a *Semantic Network* is called up (Minio & Tasso, 1996), the structure of which has a central node representing a potential topic of user interest and a set of satellite nodes representing keywords which co-occur in the same document. In this case the unknown keyword *k* is inserted as a "co-keyword" in the model and, by using the weighted semantic links present in the Semantic Network, *k* is connected to the model components found in the document.

This enables the system to distinguish between different meanings of a word by the context in which it occurs, hence dynamically widening the semantic potential of the user model and permitting far more accurate filtering. These features endow Information Filtering based on User Modeling with the capabilities of *behaviour-based* interface agents (Lieberman, 1995). Rather than relying on a pre-programmed knowledge representation structure only (stereotypes, rules, neural networks, etc.), the knowledge about the domain is incrementally acquired as a result of inferences from the user's information requirements.

BEST COPY AVAILABLE

5. Experimental Results

We have carried out some preliminary experiments to evaluate how satisfied users are with our filtering system. Results obtained with two different kinds of user are presented: one was interested in documents concerning DBMS and the other in information on Artificial Intelligence in general. Each of the users input 30 queries and then assigned a relevance value to the three documents which the system presented as most relevant. After each query, we measured the average position of these three documents (*pos_doc1*, *pos_doc2*, *pos_doc3*) in order to compare the performance of the system (rank-ordered list of the retrieved documents) with the rank-ordered list provided by AltaVista. The normalised measure adopted is:

$$Performance = 1 - \frac{\frac{(pos\_doc1 + pos\_doc2 + pos\_doc3)}{3} - 2}{total\_documents}$$

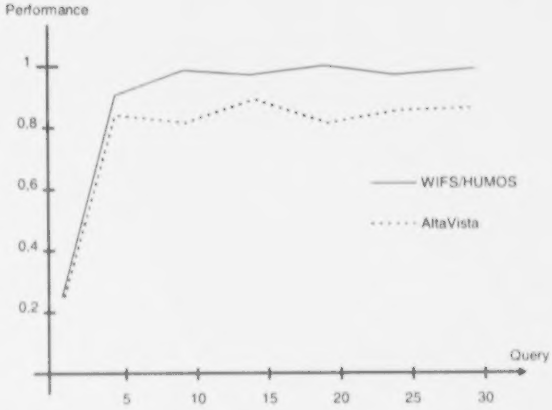


Figure 5: The average position of the three most relevant documents.

Its evolution in both cases is shown in Figure 5. As one can see, our system improves the capabilities of AltaVista by about 20% (with respect to the proposed measure of performance). The experiments conducted, although preliminary, are encouraging. The system is indeed able to learn from and adapt to users in order to deliver to them highly relevant information with a high level of performance.

References

Belkin, N. J. and Croft, W. B. (1992). "Information Filtering and Information Retrieval: Two Sides of the Same Coin?". Communication of the ACM, **35**(12), pp. 29-38.

Brajnik, G. and Tasso, C. (1994). "A shell for developing non-monotonic user modeling systems". *International Journal of Human-Computer Studies*, **40**, pp. 31-62.

Forbus, K. D. and De Kleer, J. (1993). *Building Problem Solvers*. The MIT Press.

Doyle, J. (1979). "A truth maintenance system". *Artificial Intelligence*, **12**, pp. 231-272.

Fu, L.M. (1994). *Neural Networks in Computer Intelligence*. McGraw-Hill.

Goonatilake, S. and Khebbal, S., editors (1995). *Intelligent Hybrid Systems*. John Wiley & Sons.

Kay, J. and Kummerfeld, B. (eds.) (1996). "Workshop on User Modeling for Information Filtering on the World Wide Web". *Printings of the Workshop on User Modeling for I.F. on the WWW of the Fifth International Conference on User Modeling UM-96*, Kailua-Kona, Hawaii.

Kirkpatrick, S., Gelatt, C.D. and Vecchi, M.P. (1983). "Optimization by simulated annealing". *Science*, **220**, pp. 671-680.

Kolodner, J. (1993). *Case-Based Reasoning*. San Mateo, Calif., Morgan Kaufmann.

Lieberman, H. (1995). "Letizia: An Agent that Assists WEB Browsing". In *Proc. of IJCAI 95*, Montreal, Canada, pp. 924-929.

Micarelli, A. and Sciarrone, F. (1996a). "A Case-Based Toolbox for Guided Hypermedia Navigation". In *Proc. of the Fifth International Conference on User Modeling UM-96*, Kailua-Kona, Hawaii, pp. 129-136.

Micarelli, A. and Sciarrone, F. (1996b). "A Case-Based System for Adaptive Hypermedia Navigation". In: I. Smith and B. Faltings (eds.) *Advances in Case-Based Reasoning: Proceedings of the Third European Workshop EWCBR-96*, Lecture Notes in Artificial Intelligence, **1168**, Springer-Verlag, Berlin, pp. 266-279.

Minio, M. and Tasso, C. (1996). "User Modeling for Information Filtering on INTERNET Services: Exploiting an Extended Version of the UMT Shell". In *Printings of the Workshop on User Modeling for I.F. on the WWW of the Fifth International Conference on User Modeling UM-96*, Kailua-Kona, Hawaii, pp. 9-14.

Papagni, M., Cirillo, V. and Micarelli, A. (1997). A Hybrid Architecture for a User-Adapted Training System. To appear in *Proceedings of the Fifth German Workshop on Case-Based Reasoning*, Bad Honnef, Germany.

Rich, E. (1983). "Users are individuals: individualizing user models". *International Journal of Man-Machine Studies*, **18**, pp. 199-214.

Rumelhart, D.E. and McClelland, J.L., editors (1986). *Parallel Distributed Processing*. MIT Press, Cambridge, Massachusetts.

Techniques For Enhancing Web-Based Education

Kathy Barbieri
Cornell Information Technologies, Cornell University
Ithaca, NY USA, kab23@cornell.edu

Susan Mehringer
Cornell Theory Center, Cornell University
Ithaca, NY USA, susan@tc.cornell.edu

Abstract: The Virtual Workshop is a Web-based set of modules on high performance computing. This approach reaches a large audience, leverages staff effort, and poses challenges for developing interesting presentation techniques. This paper describes the evaluation of seven techniques from the point of view of the participants and the staff developers.

Introduction

Asynchronous learning over the Web is an exciting and rapidly changing area. Cornell Theory Center (CTC) entered this arena in 1995 with its first offering of the Virtual Workshop (VW). Since then, we have expanded the scope of topics and introduced new Web-based features to enhance the learning environment. Seven techniques are presented with their technical implementation, application in the workshop, and assessment by the audience and staff.

CTC Education and the Virtual Workshop

Over the past ten years, CTC has been a leader in developing and delivering education in high performance computing to a national base of researchers, faculty and students. CTC education programs include on-site workshops, undergraduate programs, education accounts for college courses, and programs geared to K-12.

By 1995, the tools were available to pursue offering a remote workshop via the World Wide Web. Web-based materials are well suited to our education environment, which requires frequent material updates to keep pace with rapidly changing technology. We transformed our extensive set of workshop lectures, on-line tutorials, and lab exercises into Web-based modules. We began by offering a beta-test Virtual Workshop in the spring of 1995. Response to this initial test was so enthusiastic we were encouraged to further develop this format. Since then, we have offered five production Virtual Workshops to about 800 participants. These workshops covered six topics comprised of over thirty modules; Parallel Programming, Message Passing Interface (MPI), High Performance Fortran (HPF), Parallel Virtual Machine (PVM), Performance, and Scientific Visualization.

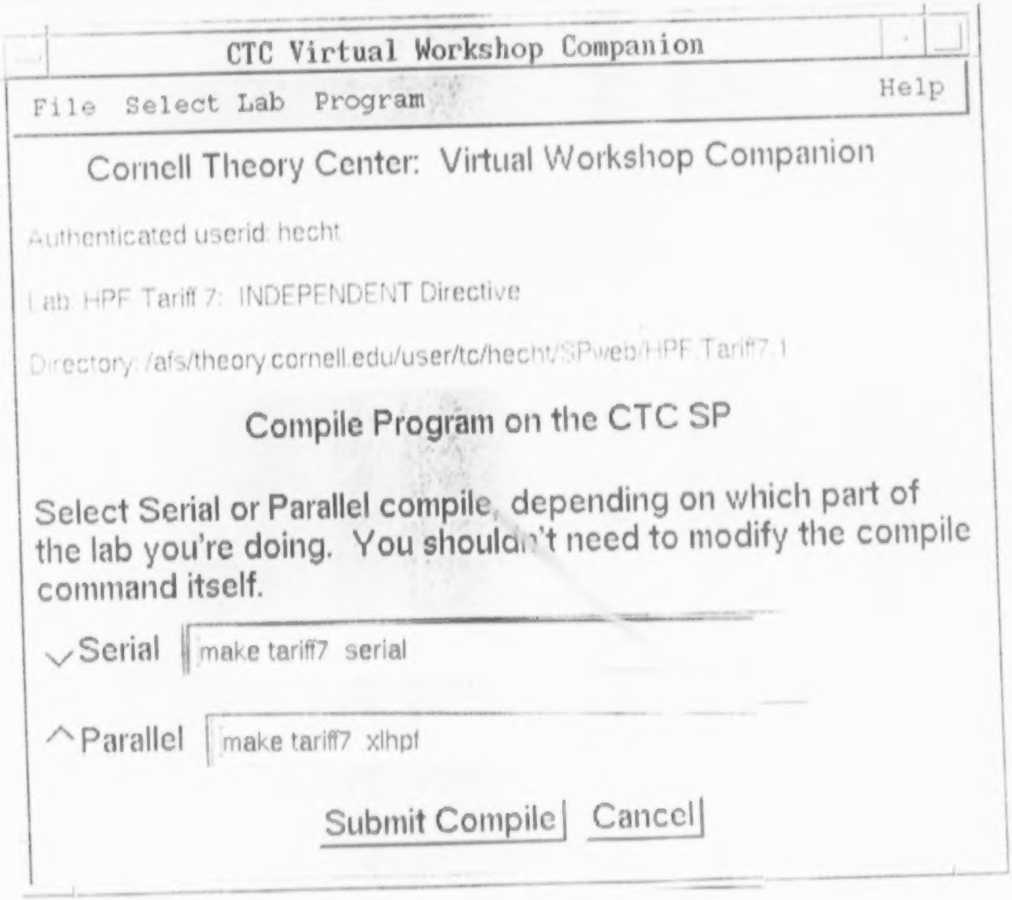
A common theme in the VW evolution is experimentation with techniques intended to provide more interaction and options for participants with different learning styles. Most of the features discussed below were designed for, or used by, an HPF module created in 1996 by the authors and others.

Technical Features and Their Applications

We designed the HPF module specifically to be offered over the WWW. This prompted us to explore methods and techniques which replace some of the interactive nature of a face to face workshop. This module builds on successful features from previous VWs and adds some new ones described below.

Java and Perl Scripts - Web-based Editing and Program Submission

A more comprehensive total learning environment has been achieved by adding the ability to edit, compile, set run-time parameters such as number of processors, and submit a program, all without leaving the Web browser. Results of the compile and program execution are displayed in a Java program window. This automatic Web-based program submission interface, called the VW Companion, utilizes a combination of Java programs and Perl scripts. The Java programs handle the front end user interface, while the Perl scripts interact with AFS Kerberos authentication to provide the required security for the users.



Cornell Theory Center: Virtual Workshop Companion

Authenticated userid: hecht

Lab: HPF Tariff 7: INDEPENDENT Directive

Directory: /afs/theory.cornell.edu/user/tc/hecht/SPweb/HPF.Tariff7.1

Compile Program on the CTC SP

Select Serial or Parallel compile, depending on which part of the lab you're doing. You shouldn't need to modify the compile command itself.

Serial: make tariff7 serial

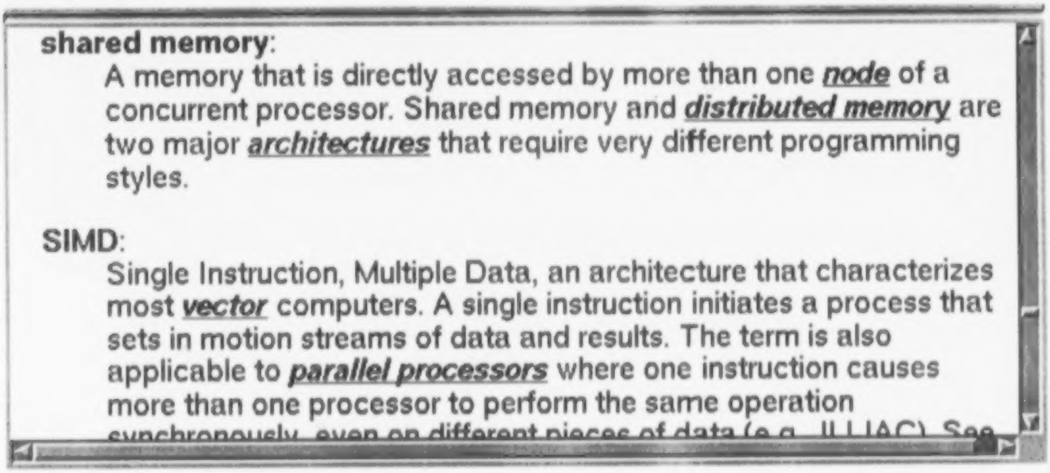
Parallel: make tariff7 xlhpf

Figure 1: Web-based Editing and Program Submission

JavaScript - Glossary

BEST COPY AVAILABLE

JavaScript was used to create a self-referencing glossary. Glossary terms in the text of the module are in italicized, bold font. Clicking on a term causes the glossary to appear in a smaller window, with that term and the definition at the top of that window. Definitions in the glossary also contain other linked terms. Implementation of the JavaScript was fairly easy. The difficult aspect is achieving similar appearance and functionality across browsers and platforms. The glossary concept is well received and seems to be 'expected' by the participants. Once the initial glossary was in place, it was fairly easy to add terms to it and use it across all our education materials.



shared memory:

A memory that is directly accessed by more than one ***node*** of a concurrent processor. Shared memory and ***distributed memory*** are two major ***architectures*** that require very different programming styles.

SIMD:

Single Instruction, Multiple Data, an architecture that characterizes most ***vector*** computers. A single instruction initiates a process that sets in motion streams of data and results. The term is also applicable to ***parallel processors*** where one instruction causes more than one processor to perform the same operation synchronously, even on different pieces of data (e.g., ILLIAC). See

Figure 2: Glossary

CGI Scripts - Interactive Quizzes

We have used CGI scripts to write interactive quizzes for our training materials. The quizzes are written as forms in a multiple choice format. Filling out and submitting the quiz form automatically grades the quiz and returns a list of questions that were answered incorrectly. An option button allows the participant to choose whether they wish to receive a detailed explanation of all quiz answers along with the grading results. Quizzes consist of a CGI script, which handles the grading, a form which contains the questions, and an ascii file listing the correct answers. The CGI script was written to handle any multiple choice quiz. Developers find it very easy to create quizzes, since only a form with questions and an ascii file of correct answers must be written. Workshop participants like the simple format and immediate answers, as well as being able to test their understanding.

Netscape Frames - Personalizing Navigation

Use of frames allows workshop participants to move through material in a way that best suits their learning style and needs. The HPF module was designed around an HPF program; workshop participants can learn the material either by working through the program or by working through the topics as displayed in the Table of Contents in the left-most frame. Frames allow coordination of material shown in two frames, by use of a simple link. In the HPF module, the reader can read about a program directive in one frame, and with the targeted link, move the program shown in another window to display the directive as used in the program. This method of organizing materials is easily adopted by staff.

Gifmerge and QuickTime - Animations

We have used gifmerge to create small, simple animations, such as this animation of an array shift. Gifmerge is popular with developers, because it is freely available via the Web and very simple to use. Gifmerge can take gifs created by any means and merge them into an animation. VW participants found that these animations enhanced the text description of the topic. Gifmerge is easily used because it doesn't require disk space or special hardware or software.

A QuickTime movie was created to demonstrate and compare speedup achieved through MPI and HPF. The technique demonstrates a successful approach of teaming the content expert with an animation specialist and creating a high quality movie which adds to the understanding of the results. It allows the author to zoom sections of the graph and emphasize the critical portions in conjunction with the narration. The movie, which runs on PCs, Macs, and UNIX platforms, was divided into three parts ranging in size from 2 Mbytes to 6 Mbytes. Based on review from our academic affiliates, we learned that downloading files this size over the internet was too time consuming and few waited for the results. As an alternative to the movie, we provided the same information in a series of graphs and corresponding text, taken from the narration. This proved to be an effective approach for the participants and required significantly less staff time and effort. As compression techniques and internet bandwidth improve, QuickTime movies will become more viable.

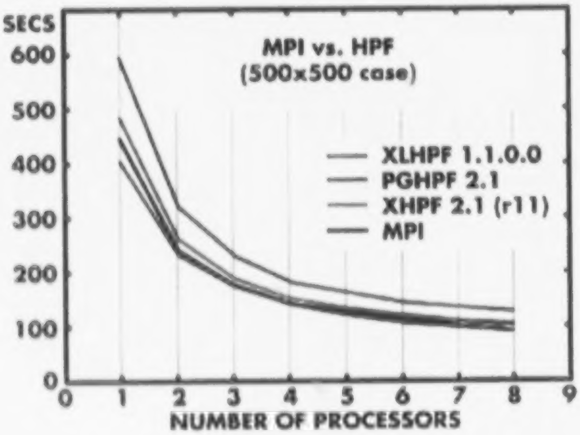


Figure 3: QuickTime Movie

HTML Link Usage - Lab Exercise Design

It is difficult to design lab exercises which target a broad audience with varying levels of knowledge on the topic. In addition, VW participants want lab exercises to be simple, yet meaningful. We have attempted to provide this by writing a lab exercise which is broken into small steps. Each incremental step is presented with a standard set of helpful links, including detailed instructions, common errors, and the solution for each step. This approach allows individuals to choose the amount of guidance they receive at each stage. A recent participant notes "The lab was quite good, since it took things in small steps which is essential to feel comfortable with parallelizing (the program) without getting overwhelmed."

Audio - Hearing From the Experts

In an effort to diversify and increase comprehension of the module, we introduced a set of audio-tagged foils. We videotaped a lecture at CTC, transferred the digitized audio to a file, edited and converted it to aiff (Mac) format and .au files. These files were then 'tagged' to the foil used in the presentation and inserted in the module. The participants were able to click on a sound icon and hear portions of the lecture in addition to reading the foil. Early feedback indicated a preference for text rather than wait to download even small file (.3 Mbytes). The staff effort involved here did not warrant pursuing this approach. Streaming audio offers a more promising approach to enhancing text.

MOO/Chat - Discussion Forums

The first two VWs offered a MOO as a forum to promote discussion among the participants as well as with the CTC staff. We offered 'rooms' in the MOO devoted to individual topics as well as scheduled times for staff to be in attendance. Few of the participants took advantage of the MOO. It required the audience to use a new tool and the scheduled times were contrary to the asynchronous nature of the VW. We then pursued a Web-based chat room which was more intuitive to use. Again we met with limited success. These approaches were not a good match for the format of this workshop. We would like to pursue more collaborative forums in the future.

Conclusions and Futures

Web-based education is an effective means for CTC to leverage its education efforts in reaching a diverse national audience. Based on evaluations from the participants, this asynchronous approach to learning affords them the opportunity to learn high performance computing without the inconvenience or expense of traveling to CTC. Convenient Web-based editing and submission of programs is in line with the look and feel of the materials. Self-assessment through interactive quizzes using CGI scripts are very well received. We have successfully used new techniques such as Java and JavaScript to enhance the interactive nature of the workshop. Network bandwidth restrictions reduce the effectiveness of some audio and video features.

Summary of URLs

Cornell Theory Center
http://www.tc.cornell.edu/

CTC education programs
http://www.tc.cornell.edu/Edu/

CTC on-site workshops
http://www.tc.cornell.edu/Edu/Workshops/

CTC undergraduate programs
http://www.tc.cornell.edu/Edu/SPUR/

CTC education accounts
http://www.tc.cornell.edu/Edu/CTC/

CTC programs geared to K-12
http://www.tc.cornell.edu/Edu/CTC/EduK-12.html

Example of an HPF module:
http://www.tc.cornell.edu/Edu/Talks/HPF/Intro/

VW Companion top page:
http://arms.tc.cornell.edu/VWCompanion/

Example of an interactive quiz:
http://www.tc.cornell.edu/Edu/Talks/HPF/Intro/distribute.html#Quiz

Example of an animation using gifmerge:
http://www.tc.cornell.edu/Edu/Talks/HPF/Intro/cshift1.html#General

Gifmerge home page:
http://err.ethz.ch/~kiwi/GIFMerge/

HPF exercise:
http://www/Edu/Tutor/HPF/Essentials/Karp/

References

[Mehringer and Barbieri, 1996]. Mehringer, S.H., & Barbieri, K.A. (1996). A Workshop at Your Workstation *Forefronts* 11 (4), Fall 1996, 10-11.

Acknowledgements

We would like to gratefully acknowledge our colleagues who have helped develop the VW, especially Anne Trefethen, Associate Director Computational Support, and Ros Leibensperger, VW Project Leader. In addition we thank those involved in the development of the features described above: Donna Bergmark, Doug Elias, Caroline Hecht, Brian Lewis, Dave Lifka, Chris Pelkie, Nikki Reynolds, and Catherine Starkey.

We appreciate the collaborative efforts of Northeast Parallel Architecture Center (NPAC) in developing the Web-based submission and the audio tagged foils, particularly Geoffrey Fox, Kivanc Dincer and Kim Mills. We thank Cherri Pancake, Oregon State University, who provided expert advice on the usability of the HPF module.

Author Biographies

Kathy Barbieri is currently Coordinator for Training and Integration for Cornell's Project 2000. Previously, during her ten years at CTC, she was Coordinator for Distributed Education, involved in conceiving the concept of the VW as well as developing some of the materials and interactive techniques. She also worked with CTC partners to bring the VW to other locations.

Susan Mehringer is a Senior Technical Consultant at the Cornell Theory Center, where she has been since 1987. Her current projects include work on the Virtual Workshop project, including developing modules, consulting, and assessing WWW statistics; testing new systems and software; consulting; and developing and delivering training materials.

All brand names and product names referred to in this paper are trademarks or registered trademarks of their respective holders.

ClassACT: The Evolution of a Collaborative Learning Tool

Richard Barone
Learning Technologies Group, Academic Technologies
Northwestern University
email: ibrich@nwu.edu

William Parod, Jr
Emerging Technologies Group, Academic Technologies
Northwestern University
email: bill-parod@nwu.edu

Abstract: This is a reporting on the evolution of a software environment called ClassACT, Class Annotation and Collaboration Tool, from a simple multimedia annotation program to a multi-domain archival database management system. ClassACT was developed at Northwestern University for instructional use and although its original goal was to solve a specific problem for a single instructor it has growth in breath of functionality (as is the norm for today's software applications) and gathered a following of users and supporters from varied disciplines. This paper will give an overview of the utility of this tool in the context of its evolution over four years in Northwestern classrooms. It will in turn contemplate ClassACT's future. A future that is at the same time promising from a pedagogical standpoint and perplexing to its developers.

Background

The teacher is one who makes two ideas grow where only one grew before.

--Elbert Hubbard

Since we a talking about evolution it is important to know something about ClassACT's history. Perhaps the most interesting comparison of then and now comes with a look at the motivation for the initial project. The project started late in 1992 as series of discussions between the Learning Technology Group and Professor Carl Smith of the Department of English and The Program for American Culture. Professor Smith's initial need was to do something he had not been able to accomplish to his satisfaction in his 20 years of teaching at Northwestern: provide students with adequate access to his slide collection. Smith was also interested in incorporating materials from the American Memory Project (then distributed only on laserdisc). A proposal to start work on the project was sent to the director of the Academic Computing and Network Technologies in early 1993.

Excerpts from the Original Proposal (February 2, 1993)

Professor Smith's goals, specific but potentially far reaching:

Carl Smith's basic goal is "to explore the possibilities of multi-media resources in humanities education". He wishes to learn how these resources can be used to supplement print materials and what it would take for faculty and students to use them in both individual and group situations.

...

The American Memory Project is a superb example of how technology has improved access to primary source materials and yet its utility on campus remains essentially unexplored. Carl envisions three ways that it can be adapted to help train students in the analysis of primary sources:

1) by encouraging "students to use the American Memory Project directly";
2) by using "the American Memory Project and other digitized teaching materials to develop new and superior kinds of classroom presentations and assignments";
3) by providing "a framework in which students can use these new resources to put together their own papers and presentations".

This next point proved to be a precursor of what would drive future forms of the project:

To accomplish this he wishes to collaborate with ITG to make appropriate portions of the AMP collection easily accessible in an "intellectually challenging and engaging form" and to build templates that would give students a new format for organizing, presenting, and analyzing these materials.

The implications of Professor Smith's proposal reach far beyond the potential benefits to his classes. Firstly, and I cannot understate this, that new technology requires new methods for it to reach its potential, and it is time for that potential to be realized. Committing to curriculum development and to a complete integration plan for existing courses, this effort aims to take full advantage of the resources and tools it embraces.

A hypothetical statement that proved accurate 3 years later:

Efforts such as this "will provide a broader evaluation of the assets and liabilities of multimedia in humanities education. There is no way to conduct such an evaluation without undertaking a project of this kind."

First Offering

In the fall semester of 1993 "The Cultural Imagination of Turn-of-the Century America", a multi-disciplinary course in English and American History, reached the classroom. The instructor's main goal at that time was to take advantage of electronic resources, especially digital image collections, to increase student access to source materials. "Cultural Imagination" established a learning environment that interconnected the literature, images (paintings, illustrations, movies), and sounds of the early twentieth century. Students not only accessed sources electronically but were also encouraged to complete their assignments, when appropriate, in electronic form. From 1993 through 1995 access to the imagebase (image database) was provided by HyperCard stacks served over Appleshare. This was a satisfactory network solution for a course with 20 students served by well equipped Macintosh labs, but soon other issues needed to be addressed. Access for Windows users, so that any student using the imagebase could access it from NU's newly wired dormitories, was imperative. Copyright agreements had to be protected. If the imagebase was to become a more generic tool and if more faculty were to find it useful, it had to provide access to external archives. It also needed to scale up from hundreds to thousands of images while preserving its flexibility and ease of use. All of these issues were addressed with the conversion of the "Cultural Imagination" imagebase to a web-based system in early 1996. This was made possible in great part by the rapid advances in WWW software tools and an institutional license with Oracle Corporation.

The Emergence of ClassACT

Professor Smith's imagebase was migrated to the world wide web for the Spring quarter of 1996. The migration required a significant file conversion undertaking that was accomplished with a commercial batch image processing utility, but the core of the effort was embodied in three major tasks:

1. An update of Smith's annotated notebooks, a major component in the course structure.
2. The development of a web-based interface for the course.

3. The development of a cgi, common gateway interface, between our server and the Oracle database that now controlled access to all media elements.

The first two tasks were mostly the responsibility of Professor Smith, which he accomplished with the appropriate technical support. The third task was accomplished by our technical staff in close consultation with Smith in order to satisfy his course model. It is this third task that we will focus on now for the server-cgi-database composite that resulted is the nexus of ClassACT.

So What is ClassACT?

ClassACT is a hypermedia document management system, a searchable media database, a groupware software application tailored for class based projects.

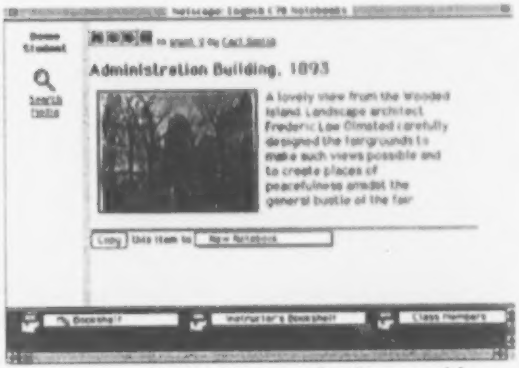
How does ClassACT apply to the Classroom?

ClassACT provides annotation, collaboration and archiving tools that allow an instructor to create an interactive learning environment. The ClassACT information web page offers:

ClassACT allows an instructor to assemble an on-line collection of multimedia, called a "notebook", for use in a course, and to provide commentary (or annotations) with each of the media in the notebook. ...Students enrolled in a course using ClassACT have access to the instructor's on-line notebook, and they may create their own versions of the Notebook during the quarter. Students have the choice to keep their own notebooks private (e.g., for self- study) or they may elect to publish their notebooks for the purpose of collaboration or for submission as a formal class assignment.

The Basic Notebook Display

ClassACT notebooks are displayed as hypermedia web pages. The default display model, the one used in Carl Smith's prototype, is that of an annotated image: a page with a thumbnail image and a block of text, the annotation. The thumbnail image links the notebook to more information about the annotation. In Smith's prototype model this was usually a high resolution version of the thumbnail image. But the link can also point to a quicktime movie file, an audio file, a text document or a local or remote URL.


Administration Building, 1893

A lovely view from the Wooded Island. Landscape architect Frederic Law Olmsted carefully designed the fairgrounds to make such views possible and to create places of peacefulness amidst the general bustle of the fair

Figure 1: A notebook page with thumbnail image and annotation area.

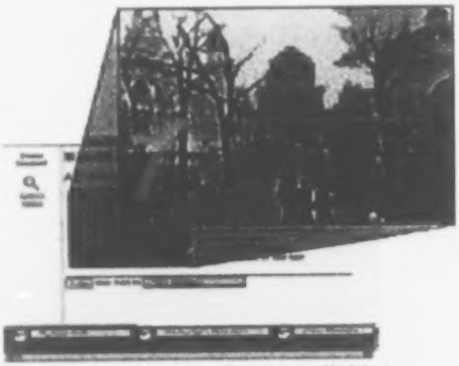


Figure 2: Smith's Prototype linking thumbnail image to Hi-Res version.

BEST COPY AVAILABLE

A frames-capable browser is required to view the notebooks. The notebook pages consists of two frames, one for the notebook body and a navigation frame at the bottom which is always available, providing access to all of the reader's notebooks, the instructors notebooks and notebooks made viewable by other class members. The navigation frame also provides access to profiles for each member of the class.

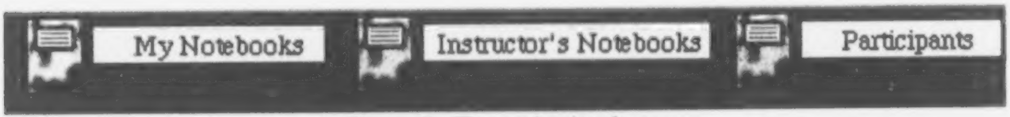


Figure 3: The navigation frame

The notebook body, in addition to the thumbnail image and annotation block, contains a set of icons in the left margin. Each icon represents a tool that adds functionality to the to current page. This format is consistent for all notebook views although the specific tools available in the margin depends on the access level granted to the viewer.

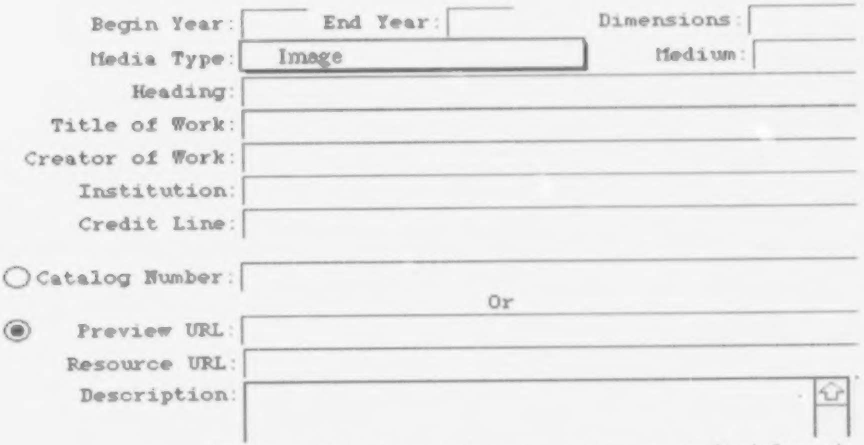
The Student Viewpoint

Students create their on-line notebooks using ClassACT's HTML FORMS interface. This is done completely within the structure of ClassACT. They do not need an off-line editor; absolutely no knowledge of HTML is required. A student has the ability to create and delete notebooks and determine who in the class can view them. They can search the media catalog, copy information from the instructor's notebooks, trade information between their own notebooks, add links from external web sites and insert their own original text.

The Instructor Viewpoint

The instructor can do what a student can with the added ability to "publish" a notebook. All published instructor notebooks are available to class members from the "Instructor's Bookshelf", which is available in the navigation frame at the bottom of all notebook display pages.

The Librarian



Begin Year: End Year: Dimensions:
Media Type: Image Medium:
Heading:
Title of Work:
Creator of Work:
Institution:
Credit Line:
○ Catalog Number:
Or
● Preview URL:
Resource URL:
Description:

Figure 4: Librarian privileges grant access to catalog information.

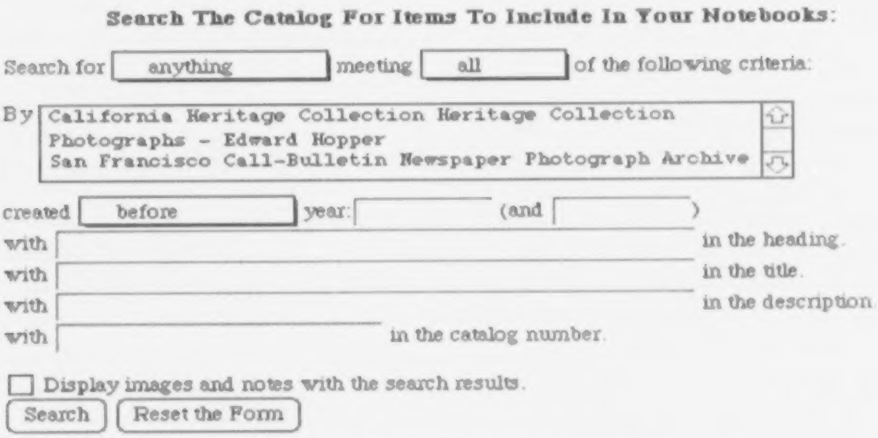
BEST COPY AVAILABLE

Instructors like Professor Smith who wish to provide access to special image collections need to first digitize the collection and then enter each item of the collection into a ClassACT catalog. In order to do this the instructor must have "librarian privileges" in addition to normal instructor status. Entries to the catalog need not be limited to local collections of media however. Anything one locates on the web, i.e. anything with a valid URL, may be incorporated into a catalog. In this respect the catalog can be used as a sophisticated bookmark manager.

The Administrators (User, Domain and Group)

The prototype for ClassACT dealt with user administration: a system for adding and deleting members from the class. As ClassACT became available for multiple classes there needed to be a way to separate the various functions of ClassACT for concurrently active classes. The solution for this was to implement the concept of domains. Each class can now be assigned its own domain with independent catalog, class roster and notebook collections. The domain administration privilege is required to create and manage a domain. A group administration feature was also added after the prototype. A group administrator can define workgroups of two or more students within a domain (class). Students in the same workgroup have the option of exclusively sharing a notebook within the group.

Searching



Search The Catalog For Items To Include In Your Notebooks:

Search for anything meeting all of the following criteria:

By California Heritage Collection Heritage Collection
Photographs - Edward Hopper
San Francisco Call-Bulletin Newspaper Photograph Archive

created before year: (and)

with in the heading.

with in the title.

with in the description.

with in the catalog number.

☐ Display images and notes with the search results.

Figure 5: ClassACT search options.

Since an Oracle database administers the notebooks and the media catalog, its extensive search capabilities are at ClassACT's disposal. Searching is nonetheless limited to the information recorded for catalog items. Since simple searches were adequate for the prototype course, catalog detail was a low priority. However, recent development of the product, looking to expand the utility of ClassACT, has concentrated on defining a catalog record that is consistent with current archival digital media projects. Down the road it may be possible to create and maintain an extensive legitimate shared archive of digital media with the ClassACT interface.

The Big Picture

A Summary of Results and Benefits

The man who can make hard things easy is the educator.

--Ralph Waldo Emerson

The recurring theme of the development team from the original HyperCard version of Professor Smith's class to the media notebook model of ClassACT was that "it has to be easy for the students to use". Smith above all else insisted on this. With this understood the general objectives and priorities were defined and remained essentially the same throughout the entire evolution of the project:

1. To provide students with extended access to source materials of various forms.
2. To improve student-teacher and student-student communications.
3. To stray from traditional course structures by providing tools that would allow students to electronically publish course assignments and journals.

Instructor and Students Benefits

It is safe to say that at least for the Smith class model, which will be taught for the fourth time in the fall '97 quarter, all of these objectives were met. According to Smith the instructor-student benefits were significant. The increased opportunity to analyze source images resulted in a broader participation in class discussions. The electronic paper was made available as an optional form of expression in his course. There was no contention, nor was there any expectation, that an electronically delivered paper would be superior to a printed one. Electronic notebooks offered their authors the ability to link directly to a source image in contrast to making a written reference. Most of the students throughout the history of the Smith model chose to submit at least one assignment as an electronic notebook. Those students who chose to produce class projects as electronic documents rather than traditional printed papers, according to Smith, "did as well".

Developer/Support Benefits

Converting the original "mediabase" project into a web tool has provided cross-platform access and expanded the scope of the tool by adding the ability to link to other appropriate web sites. The implementation of a full function database made the application scalable. Improved searching capabilities and the "potential" to cope with large classes has made ClassACT attractive to a larger segment of the faculty at Northwestern. The Oracle database component also adds security for access to a ClassACT domain now requires an account name and password. This is an important feature when the issues of copyright and fair use are discussed with faculty.

The Challenge: Where are we going with this?

Just how will ClassACT scale up? Can we handle a large class of 200 students that could generate many simultaneous hits to the site? Are our accounting and document management components robust enough to handle more than two or three separate classes? Certainly the Oracle database has been used in far more demanding situations but will we have to continually modify and update the cgi scripts that form the underpinnings of ClassACT? Should we convert the interface from Pearl to JAVA or C++ to speed up what may be a bottleneck as we accept larger numbers of users? Should we continue development in-house or should we look to commercial software tools to handle some of the tasks, such as document management? Should we just freeze development here and apply what we have learned when we go shopping for commercial software? Should we seek a collaborative effort with other universities, pooling resources, to continue and refine the product?

If ClassACT doesn't capture the imagination of faculty and provide instructional solutions for a significant number of them then these questions will never have to be acted upon. This is however not the case. ClassACT has been meet with enthusiastic responses whenever we have demonstrated it. It is currently being used in one Art History class at

Northwestern and under consideration for two others, one a large lecture hall class that would allow us to test one of the above concerns. Instructors at other universities have also expressed interest in using the tool.

But the most challenging new application for ClassACT will come in the fall when it is used in a Political Science course that deals with the effects of media on government policy. Student's in the class will be required to collect and display media for assignments. This means that students will require "Librarian privileges" so that they can add their digitized media files to the domain catalog for the class. In all previous application of ClassACT only the instructor and teaching assistants were allowed to modify the catalog for a class. This will be some work for us but it is the kind of innovative use of technology that we seek to support.

Still the biggest challenge lies with the instructor who is determined that all class assignments will be created within the framework of ClassACT notebooks. This "no print option" is a most significant commitment to the technology. In previous offerings of this class the instructor has had difficulty gathering the various elements of her students multimedia projects. She found it equally difficult to get feedback to them. She sees ClassACT as an appropriate organizational tool and a great time saver for her and her students. Where Professor Smith sought to offer an additional form of expression with his notebook model, this instructor intends to completely replace what was for her an inefficient traditional delivery method with a more appropriate one. Maintaining the traditional method for her would be like finding a solution but keeping the problem.

Two Models Surface

These recent adaptations highlight what we have for sometime suspected: ClassACT addresses two diverging application models: the shared archive model and the cognitive process model. The large Art History course which requires access to an extensive collection of media points in the direction of the former. Courses like this rely heavily on the cataloging capabilities of ClassACT. Notebooks are used as paths through a growing maze of media information.

The Political Science media course, with it's heavy emphasis on using media in the learning process is the epitome of the later model. This model demands a high degree of interactivity and special attention to improvements to the computer-human interface in future versions Although some class applications may require components of both models, as in the Smith prototype, there are enough special needs in each to suggest independent development.

Which model will prevail ?

Are we really developing two tools? The "shared archive model" is probably the easiest sell to faculty. Most of the work required of an instructor in this model involves cataloging media and creation of instructor notebooks. Since this should be completed before the first class meeting, virtually no attention to ClassACT is required of the instructor when the class is in session. The early adopters of the "cognitive process model" are entering unexplored territory and the amount of effort they will have to commit to will be harder to gauge. Members of this group of faculty will likely place the greatest stress on ClassACT often requiring new features as they ultimately see new possibilities.

It might be easier to concentrate on one model and try to learn as much as possible from it. Yet from our viewpoint as instructional support staff, both models are welcome. The most important component in any educational undertaking is good content. Once the content is defined by the instructor, choosing the "appropriate" technology becomes the focus of implementation. Our experience has been that both models provide instructors with a means of delivering and enhancing their content although under different circumstances. We have grabbed the tiger by the tail; do we hold on with two hands?

Authoring and Development in an Online Environment: Web-based Instruction Using SGML

Paul Beam, The Department of English, The University of Waterloo, Canada,
pdbeam@watarts.uwaterloo.ca

Abstract: This paper describes the research and development over two years of a fully Web-based credit course at the University of Waterloo. 100 students from across Canada complete four exercises in technical writing over a four month term. In the process they come to understand and apply the principles of SGML design in their assignments and in the profession generally. Their assignments are converted and displayed on the course web site, (at url: http://itrc.uwaterloo.ca/~engl210e/) for markers' assessments and for their shared use. Students employ the same tools, software and designs we use to create all aspects of the course materials. Through this exposure they come to evaluate our instructional methods by practicing them in their own work. Half the course grade is derived from students working entirely online in groups of three. All aspects of instruction are supervised and integrated by instructors using newsgroups, chat, online tutorials and instructor comments. We are now making a commercial version of the course available to organizations for their training needs.

Introduction: the Overview

Because this course is extensive in its content, learning options and scope of application, I wish to examine it under these categories. An overview of its **technical structure** permits the reader a view of both our **conceptual notions** of its use and the **administrative controls** necessary to operate it. We can provide something of the course from a **student perspective**. Our **research** from the outset, in the choice of an SGML learning model, has influenced how we make the course available and how we encourage participants to further develop their skills and knowledge. We are now investigating subject-specific materials for uses in other academic and business applications. We have expanded our research to include how individuals learn, how groups who meet only via the Web work together and how companies can begin to use their employees as collective resources to begin complex training within their own organizations in an entirely online environment. Finally, we have begun to **commercialize** our tools and software to make packages available to partners and clients in business for them to author and offer full-scale, interactive, web-based learning in any subjects of their choice.

Technical Structure

This course, English 210G in its credit mode in the English Department at the University of Waterloo, consists of some 5, 200 files in primarily HTML format on a fileserver running Linux and Apache. It also runs on Sun UNIX machines. The course is extensive in its contents of some thirteen 'modules' of various technical writing topics in interactive formats, an SGML editor, Document Type Definitions for some six technical writing documents, SGML converters to render the SGML data into HTML and RTF formats, software to enable multiple chat sessions and links to language tools, dictionaries and remote sites for information on technical documentation. The online assignments can contain full multimedia options, executable programs and CGI scripts and forms, though some of these latter require special arrangements for conversion. Students are encouraged to learn about and attempt sophisticated expressions but the grading centre of the course remains English standards of expression of text.

The course is supervised by an instructors group consisting of the professor, the course contact personnel, technical support, markers and writing experts whom the students may consult via email from various locations within the course. A complete version of the course exists as a .tar file at our FTP site for distribution to partners in other research locations. All course materials are created and maintained as SGML master files which are converted and placed in directories from which users access them as HTML files from the course web site at the url listed above. Students create and convert their assignments by the same authoring processes we employ and this is one of the major instructional features of our pedagogy. Students submit all assignments to the course server as SGML files and they must convert completely there to HTML displays within the "View Student Assignment #" on the course Bulletin Board, at url http://itrc.uwaterloo.ca/~engl210e/BulletinBoard/ . We provide local conversion software so students can edit and test their documents prior to submission and we give what technical help we can in tutorials, FAQs and by responses to email to support a complex and often confusing process for remote users.

These student materials can be easily linked or incorporated directly into modules as illustrations and examples and advanced student projects can join the course as learning materials in their own right. We use SGML because our converters automatically update the resulting HTML files and their links and reduce web site maintenance while they guarantee complete accuracy of all links within our structure. The Course Administration Tool has been brought on line this fall to assist the instruction team in the maintenance of course accounts and the collection, conversion, distribution and return of student assignments which are administered and marked entirely in an SGML format. The instructor provides comments to the entire group, sets up and chairs chat sessions, edits newsgroups for content of general interest, maintains the Frequently Asked Questions lists and keeps specific course information current from an administrator's 'flight deck'. Inquiries on more technical issues are welcomed by Mr. Brian Cameron of the University's IST Group at email hesse@cuckoo.uwaterloo.ca.

An Instructor' View

English Technical Writing 210G is a comprehensive learning package which includes:

information about academic requirements
an introductory package at http://itrc.uwaterloo.ca/~engl210e/InformationDesk/IntroPackage/intro.htm
scheduling, and assignment submission requirements
the course administrative structures and personnel and complete class lists of names and email addresses
ftp and print options (at http://itrc.uwaterloo.ca/~engl210e/BookShelf/Introduction/CourseSoftware/)
information on technical writing standards, authorities and professional organizations
models and procedures for editing and assimilating online documents
tools and information sources for specific technical documentation.

It includes as software an SGML editor, InContext2, proprietary converters for HTML and RTF display from the SGML source documents and a series of Document Type Definitions specific to technical formal documents – the manual, report, letter and resume.

The course in its large-class expression accommodates over 100 students across Canada. Participants must have basic hardware requirements for Web access, email and a University account, rendered at time of registration. The SGML editor is available in a PC format only but we permit MacIntosh or UNIX users if they can provide their own editors. Weekly WebChats are optional but the sessions are posted on the Bulletin Board (url http://itrc.uwaterloo.ca/~engl210e/BulletinBoard/). Class members must participate in groups of three in the second and third assignments and their group mark applies to all members.

The large sections are composed roughly of 20% users from all faculties, using the campus watstar networks, 40% computer science and engineering students, many participating remotely from co-op work term placements and 40% as Distance Education students from work and home units at points across the country. Many senior students take the course as a way of becoming current with real business applications and practical sources for the reports and manuals for which they will be responsible in their career placements. Members of this latter group provide a great deal of the advanced technical knowledge manifest in some of the later assignment submissions, many of which are highly interactive and broad-ranging in their subject matter and illustrations.

Conceptualizing Online Learning

Perhaps the hardest thing for us to understand across the development of this project has been the levels of difficulty in reducing instructor involvement among the many, linked processes available to students. Common wisdom would urge that sophisticated data, complete and reliably developed, would provide a base for the confident execution of assignments clearly linked to it We have found that the course is elaborate in its design, with much more material, research capacity and personal communications features than are minimally necessary. It does, however, employ tools and processes in common use in business, so we believe our students will be expected to be familiar with these in work situations.

To account for the high degree of student and instructor involvement, we have coined the term 'richness' to describe the wide range of options for acquiring knowledge and exchanging opinions and information. In this sense, English 210G has proven to be quite untypical of both conventional classroom exchanges and of most other online learning expressions as well. Students concur that the experience is unique.

We conceptualize the model as a triad of data-communications-human interface and we have developed techniques to enhance individual learning in each sector. As well, we have integrated the sectors so parts of each link users at appropriate times to the other options. All our research effort now is directed to the transfer of human-controlled activities, with their high demands on personnel, into machine-assisted ways of providing answers to users. As example, we attempt to respond to course email as quickly as possible but we lack the resources for a full support service spread across a seven day week and five time zones. Our answer is to respond as best we can to each inquiry, but to capture the answer in a large, integrated FAQ database so subsequent questions can be directed as automatically as possible to it. The information in the base has to be linked in turn to the course data itself so users are prompted to it before they encounter the problem in communications or assignment areas. Solving these issues with enhancements to our software and modules is our most immediate task.

Student Perspectives

From the student's perspective, the course is, at first experience, a complex collection of online texts, communications devices, lists and tools for constructing and submitting SGML files to complete four course requirements in a series of progressively sophisticated technical documents. We have established the learning experience on two, linked metaphors. (Students, and my readers, can begin the demystification process with a Guided Tour at url: http://itrc.uwaterloo.ca/~engl210e/InformationDesk/engltour/tour.htm). The course Introductory Package is made available by FTP from the Home Page and is sent as floppy disks to all off-campus class members. Formal registration results in a course account for the duration of term.

The first is the course home page and depicts the floor plan of a small company specializing in technical documentation (url: http://itrc.uwaterloo.ca/~engl210e/) to which each student must apply for employment by submitting a resume and letter of proposal to join a group. This is the first of four assignments and it initiates all members to the SGML editing and conversion processes and to each other and the course instructors by reading each others' proposals in the selection of partners to commence subsequent work.

The second metaphor is a simple bookshelf from which all aspects of the course can be selected as volumes (url: http://itrc.uwaterloo.ca/~engl210e/BookShelf/). These have links among themselves to provide learning connections among assignments, instructions, tips and supporting information, as well as contacts to subject -area experts and sources of additional specialized information.

The entire process is a loosely structured narrative in which the students, as employees, work within and for the company management, the instructors, to complete a series of real technical documents, under real time constraints, within the specifications, standards and models of a real contract. The 'tale' is set in a series of rooms within an organization where members can conduct research, create and modify interactive 'documents' for submission and online display, communicate with each other, individually or in groups and use the metaphors as settings to view in-course materials, external information sources, each others' work, course tutorials and instructors' comments on group activities and their personal progress. Members must interact with the instructors and each other in professional ways, communicate and conduct business formally and produce materials to a known and objective standard. They can view the assignments of all other members (as can any visitors to the web site at url: http://itrc.uwaterloo.ca/~engl210e/BulletinBoard/). While their electronically marked assignments are password-protected, they can request clarification on grades from the markers, refer to the assignment tutorial for further comaprisons and exchange any grade information and comments with classmates. Instructors arrange special chat sessions on a range of topics, on suggestion and by request, and collect information for all course members from email, newsgroups, chats and tutorials, to present it in the daily Instructors' Comments files. Students can consult all other course members from a course list of names, email addresses and, optionally, phone numbers.

Across the four months of the course students become familiar with SGML technologies at a clerical level. Only the most advanced can actually work with SGML design and structure issues at completion. Most become comfortable with an SGML editor, come to see its advantages in permanence and document design and understand basic conversion concepts. The course provides much additional information and help beyond their assignment needs and points to areas of learning they can pursue in employment later.

We presume an involvement rate of 8 hours per week for a conscientious student seeking to participate and to complete the assignments in good order – at about a "B" level. Depending on the individual's levels of technical experience – and the course **is** called **'technical writing'** -- , more time may be required at the outset. A point to note is that a new participant has a huge 'concept ball' to incorporate. We use email, an SGML editor, chat groups, newsgroups, interactive HTML displays and a large, image-mapped database of material, parts of which have to be explored merely to get assurance that you are observing the components of the schedule to not fall behind in unknown areas. We have had to make extensive additions to our design to assure readers that they have completed sufficient exploration to commence assignment preparation. Nervous members have to be assured that this is one course of five in a term and that they can complete it successfully in the allotted time.

Typically a student might spend six hours becoming familiar with an assignment within the course site, another six reading and incorporating materials to apply to a submission, six more over several days communicating with group members and friends on organizing the structure and detail of the joint response and an indeterminate number in learning 'technical' issues. Because these vary so much across the uninitiated, the very experienced, the highly motivated and the technically well-placed (like those working in good technical writing departments, with lots of collegial advice), it is hard to generalize about 'class experiences'. During this time a good deal of linked learning is going on. Because the course is so technical and innovative, many students provide guided tours for their buddies in computer science and engineering – and we get good ideas from strange sources. Others become involved in ancillary databases like the Society for Technical Communications' web site. Some make contacts with our 'business support team' – technical writers who work with us from business situations and who advise students for volunteer service, but also to keep an eye on good recruits for their departments. These contacts have assured that we have had more available jobs than students in some six offerings of the course.

Students, across five thousand miles, form a working, academic unit in new ways. Not all participate actively in all parts of the course. Some use "chat" extensively and develop social contacts thereby. Others add to and use newsgroups, providing very specific information for particular systems issues. Some use them to develop social-pedagogic positions which they then incorporate in final reports and test in subsequent terms. We do not monitor email among individuals but we are aware of the amount of 'talking' that goes on throughout – and after – the term. Some of this percolates up to instructors as questions, ideas, suggestions, requests, at a very high level of intellectual involvement. And we try to capture this in feedback to the class via the instructors' comments, in aspects of 'Frequently Asked Questions' and in projects directed to new development and the revision of existing modules.

What the course is to students is hard to summarize. We have a majority of enthusiastic graduates, many of whom request additional learning in this mode. Drop-out rates are remarkable low, at about 5% where 25% in a course this size is common, but this may reflect the motivation and experience of members, many of whom are adults, often working from inside companies which then utilize their experiences and skills directly into online training projects using similar methods and technology. All off-campus participants must have Web access to join and the course applies a great deal of Web technology directly to learning and creation methods, so users experience a high degree of gratification. The course is immediately applicable in many of their other activities and it shows them powerful new tools, both within its own structure and at other locations. We encourage motivation in projects like comparing this course to other online learning models for credit and we often permit students to bring their own interests to assignments.

Those experiencing greatest difficulty, more as attitude than as situation, are on-campus members, some of whom never make the transition to a 'class that does not meet in a room'. Some pass the term seeking direct personal contacts for their instruction and support, unaware and uncaring that some 80% of their fellows live up to three thousand miles away from the fileserver on which materials are distributed. Some never self-motivate and for them the experience remains incomplete. Others understand and elaborate the concept of 'richness'. For these members, the added dimensions of enhanced communications, comprehensive records and the easy, efficient transfer of data make all aspects of the course desirable. They take the option of the subsequent specialized course, produce materials we then incorporate into the main course materials and often apply their new skills directly in business and government documentation departments.

Research

Our initial research lay in the development of the course itself. This consisted of determining the structures to permit a significant learning experience for a large, widely distributed group, in an entirely online format. This in turn required us to make decisions in late 1994 about the standardization and distribution of tools and hardware from our servers across the newly forming World Wide Web. A stint as Director of the Procedures and Documentation Department of a large Canadian Bank convinced me of the value of SGML as a concept and a technology by which I could provide a legacy to course members. Whatever new technologies emerge, the underlying virtues of SGML's structural requirements and its universality of expression across platforms convinced me that a) students wishing to become technical writers should understand the concepts and practices of 'mark-up' and that they should know the virtues of consistent structure within the idea of a 'document'. I also saw that these principles would continue to influence Web development, because HTML is a limited subset of SGML – the result of a more comprehensively structured document design. If we could create even rudimentary converters to render accurate, complete SGML into HTML, and also into RTF for the full range of print options, we had a very useful set of tools for students – and a very powerful set of large-document development and maintenance software for commercial use. On these research premises we built the course, primarily with student labour. We have retained those development principles and, without exception, all members of watAGE have come to the company via the course, and all remained involved in the development of aspects of its design and operation.

We now conduct formal studies into how class members learn. We commence with analyses of four, optional, anonymous online evaluations across the term, between the submission and the return of marked assignments. We also have extensive profiles of individuals because they complete quite sophisticated resumes early in the course. We work directly with them in the development of their assignments, in groups and with individuals, and many complete their final work as an evaluation of their experience in the context of their field and subject of study. We know the locations and operating systems of all members as well. As a result, we can analyze individuals' background preparation, levels of participation, sources of support, among course materials, instructors, other students and off-site resources. This work is helping us very much in the design of new course tools and has led us to the major decision to split the learning process into an initial HTML-only course, with the same designs and materials and a subsequent experience, (still using our company-bookshelf metaphors, but with participants now as managers!) of SGML to an advanced stage of document design and DTD enhancement.

We have begun to understand that a major advantage of online learning is the individual's ability to study at one's own pace and depth of insight, at the times and for the duration one requires. This process more closely resembles 'private learning, reading, the study of texts, than it does conventional classroom activities. This pattern of searching, reflection, self-testing, is set in the context of expansive communications by the individual's ability to place ideas and questions to both instructors and one's peers, in various combinations of urgency. It allows for reflective responses and the careful addressing of specific aspects of a question to trusted sources – a friend, a knowing marker, a specialist. The major benefit of the technology is the controlled lapse between question and response – the student's highly valued ability to think before having to 'speak', or to even remain silent and observe, or ask some other member of the audience for an opinion. This heightened, broadened involvement makes a wonderful atmosphere for degrees of participation and student evaluations identify and praise it.

To build on our awareness, we want now to begin a major development of both course materials, but also of the 'learning environment' for ESL members, both within Canada and abroad. We want to enhance our already extensive writing tips – the many modules on the bookshelf – with guides, diagnostic and advice tools and templated modules through which students from language backgrounds other than English can develop technical writing expressions with confidence. We perceive this to be a real need ideally responsive to the online features which provide time to reflect, to retry solutions, to seek help privately. The world needs technical writing instruction and a majority of that world needs support and assistance in English expression as well.

Commercialization and Partnerships

The course is available presently as a credit offering and as a 'continuing education' (non-credit) as 'distributed education' within the University of Waterloo. We offer versions of the course to commercial organizations and government agencies for training in specific topics and company needs. We are in the process of forming the course as two units, the one instructing in HTML-based technical writing, the other providing SGML design and operation, and we have begun development of a package of our tools and data for license by companies to produce their own SGML-based online training and large-document management facilities. These products will be available in four months, with some good planning and good luck and can be examined and applied for at url: http://watage.com . We have seen the need for

English language training and support and are beginning to work with partners in the Czech Republic and Indonesia. We regard the need for administration tools to be as critical as any authoring/creation issues, though most software developers emphasize the latter and most novice instructors see their greatest need as the ability to create data in interesting, effective ways. Our 'online instruction product' will reflect the model we have come to respect from our experiences in developing the technical writing course. It will resolve instructor difficulties as an **authoring device** by using SGML technologies to create and support linked, sophisticated, interactive modules which we convert and maintain as HTML expressions. The **communications package** will contain an easily supervised set of email, newsgroup, tutorial, chat and instructor facilities to assure simple, reliable and extended information exchange among all members. Finally, the **Course Administration Tool** permits an instructor to build and modify course content in any subject area, including the installation of executable programs in the SGML structure. The instructor will be able to incorporate members of an entire class (possibly numbering in the hundreds) into the course account structure, by individual and group entries and build directories for the reception, marking and return of assignments and projects. In effect, the triad of 'data-communications-human interface' is addressed in a comprehensive, integrated set of tools and the resulting activity is maintained and distributed across the Web for application in any academic or corporate environment.

All of our experience, in the technology, but more importantly, among its users, is that this pedagogy will develop rapidly as a favoured choice in business, in education and among motivated learners around the world. We invite you to test the model with us and to participate at your level of need.

The Freedom to Choose: Transforming Content On-demand in the BSCW Shared Workspace System

Richard Bentley[1], Torsten Täschner, Wolfgang Appelt
Institute for Applied Information Technology (GMD FIT.CSCW)
German National Research Centre for Computer Science
D-53754 Sankt Augustin, Germany
Email: {bentley, taeschner, appelt}@gmd.de

Abstract: Our studies of information sharing on the Web with the BSCW system have revealed a need for server-side content transformation tools. Specifically, we have discovered requirements for on-demand format *conversion*, *encoding* and *archiving*, which we believe to be general requirements for information sharing on the Web. We highlight these requirements using examples from our case studies of BSCW and present our solution: a transformation assistant which provides content transformation services, extensible by addition of new conversion, encoding and archiving utilities. We discuss the relevance of this work for activities in the area of HTTP content negotiation and show how a content transformation assistant, as an extension of a standard Web server, can provide users with freedom of choice without burdening information providers to make information available in different formats.

Introduction

Information providers who make information accessible via the Web determine the data formats in which information is supplied (HTML, formats of specific word processors, zip-archives etc.). Potential information consumers must have the capability to process the information in one of the formats provided or transform it to a format they can handle. The current trend is towards extending the capabilities of Web browsers (through 'plug-ins' or as part of the basic functionality) to increase the range of data formats that can be decoded, processed and displayed.

This paper discusses an alternative, complimentary approach: the provision of server-side content transformation services. We argue that extension of Web servers to provide such services is in many cases more appropriate than client-side tools, reducing the burden on administrators to install client-side software and on content providers to make information available in different formats. We use experiences from deployment and evaluation of our BSCW Shared Workspace system [Bentley et al. 97] to highlight requirements for server-side content transformation and present a solution based on a transformation 'assistant'. Finally we discuss the relation of our work to current activities in the Web standards community with respect to content negotiation in the HTTP protocol.

The Need for Content Transformation in BSCW

BSCW (Basic Support for Cooperative Work) is a project at GMD which seeks to extend the Web with basic services for collaborative information sharing. The BSCW Shared Workspace system is the basis for these services, which include features for uploading documents, event notification, group administration and more, accessible from different platforms using standard Web browsers. See [Bentley et al. 97] for a detailed description of BSCW.

Sharing Information with the BSCW System

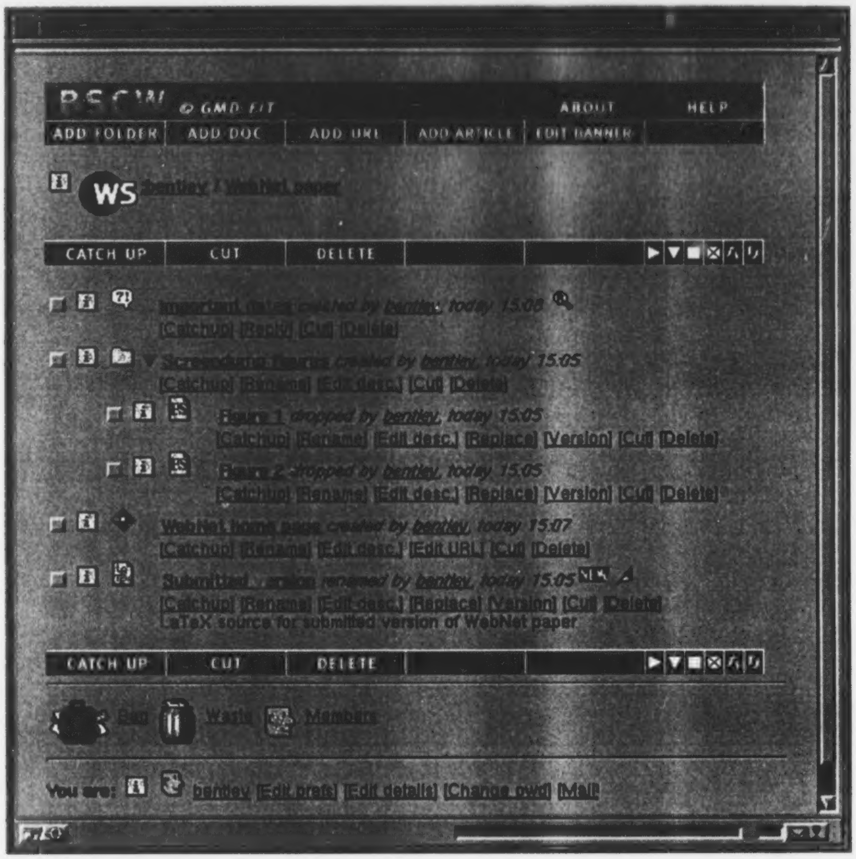
The BSCW system[2] is based on the idea of a 'shared workspace' which the members of a group establish for organising and coordinating their work. A shared workspace is a repository for information, accessible to group members using a user name/password authentication scheme. A 'BSCW server' (a Web server extended with the BSCW software through the CGI) manages a number of shared workspaces for different groups and users may be members of several workspaces, perhaps corresponding to the projects the users are currently involved with.

---

[1] Author's current address: Rank Xerox Research Centre, 61 Regent Street, Cambridge CB2 1AB, UK.

[2] All BSCW software described in this paper is available for anonymous downloading, free of charge. See http://bscw.gmd.de for details and for more information on the BSCW project.

A workspace contains different kinds of information *objects* arranged in a folder hierarchy. In [Fig. 1] for example the workspace "WebNet paper" contains an article which holds a simple text message ('important dates'), a folder containing postscript files ('Screendump figures'), an URL link ('WebNet home page') and a LaTeX file ('Submitted version'). The last 'significant' operation performed on each object is described, and a list of clickable 'event icons' give information on other recent changes. Operations which can be performed are listed below each object and a description is presented if one has been supplied (as with the LaTeX document 'Submitted paper').



:bentley / WebNet paper

☐ important dates *created by* bentley, *today 15:08*
[Catchup] [Reply] [Cut] [Delete]

☐ Screendump figures *created by* bentley, *today 15:05*
[Catchup] [Rename] [Edit desc.] [Cut] [Delete]

☐ Figure 1 *dropped by* bentley, *today 15:05*
[Catchup] [Rename] [Edit desc.] [Replace] [Version] [Cut] [Delete]

☐ Figure 2 *dropped by* bentley, *today 15:05*
[Catchup] [Rename] [Edit desc.] [Replace] [Version] [Cut] [Delete]

☐ WebNet home page *created by* bentley, *today 15:07*
[Catchup] [Rename] [Edit desc.] [Edit URL] [Cut] [Delete]

☐ Submitted version *renamed by* bentley, *today 15:05*
[Catchup] [Rename] [Edit desc.] [Replace] [Version] [Cut] [Delete]
LaTeX source for submitted version of WebNet paper

You are: bentley [Edit prefs] [Edit details] [Change pwd] [Mail]

Figure 1: A BSCW shared workspace

After an operation a new page is returned showing the current state. Besides the links below each object, buttons at the top of the page provide further operations for the current folder; 'add URL', for example, would return a form to specify the name and link of a URL object to add to 'WebNet paper'. The checkboxes to the left of each object with the buttons above and below the object list allow operations on multiple selections. Members can upload documents (we support both POST and PUT methods [Bentley et al. 97]) and set access rights. Clicking on a document downloads it while clicking a folder displays the folder's contents. This last method of navigating 'down' through the hierarchy is supplemented by the navigation bar at the top of the page presenting the current location and path; clicking on the first element of the path (":bentley" in [Fig. 1]) returns to the user's 'home folder', listing all workspaces of which the user is a member. Users also have access to their private 'Bag'; a clipboard for temporary storage or for moving objects between folders. The Bag is always accessible to the user by

BEST COPY AVAILABLE

clicking on the bag icon below the workspace listing [Fig. 1].

BSCW in Use: Requirements for Transforming Content

BSCW as described above provides no features for transforming content; successful information sharing requires use of document formats that collaborators have the necessary software to handle. We have evaluated this and other aspects of BSCW in the context of CoopWWW; a project in which BSCW forms the kernel of a more comprehensive Web-based groupware environment [Appelt 96]. CoopWWW involves cycles of development, deployment and evaluation with the project's user groups. We consider here findings from the first evaluation phase which focused on the use of the BSCW system by one user group, an organisation which provides scientific consultancy services. The organisation is distributed world-wide with offices in the UK, USA, Spain and Sweden.

Data was collected by way of questionnaires and a user workshop where problems were discussed with developers. One of the first findings was that users were having problems with data formats. Within the organisation PC, Macintosh and Unix platforms were in use and documents were being uploaded in application- and platform-specific formats which other users did not have applications (or correct versions) to handle.

To address this the organisation purchased a 'plug-in' utility for Netscape that interpreted different data formats and displayed the results in the Web browser. This required the administrators to install and update the software on all users' machines, and only allowed *viewing* documents in a workspace; *modification* still required the creator application, inhibiting collaborative document production. The organisation was then forced to establish conventions regarding 'permitted' data formats when there was a need for joint document production. These conventions also outlawed use of compression and archiving due to the lack of cross-platform availability of suitable utilities.

These conventions revealed a further set of problems, previously hidden when users could archive and compress documents before uploading. Bandwidth problems, first from the organisation's London office to a BSCW server at GMD then between satellite offices following installation of a server in London, showed that archiving and compression allowed optimal use of the channel to the server. Archiving required only one upload request to the server, while compression reduced transmission time considerably. The establishment of conventions necessary to achieve interoperability thus required more cumbersome and time-consuming upload and download procedures.

Analysis of these problems has identified the need for on-demand content transformation services:

- *Format conversion*: Conversion of workspace documents prior to download is required for users to view/edit documents for which users lack suitable application or conversion software. This would remove some of the need to install and maintain conversion and viewing software on all users' machines. Furthermore, users would have immediate access to new convertors following installation on the server.
- *Encoding/decoding*: With low bandwidth it is more efficient for users to upload [download] compressed data with the server [client] handling decompression. Conversely, it should be possible to decompress a compressed workspace document prior to download if that compression format cannot be handled by the user's machine.
- *Archiving/extraction*: Upload [download] of multiple documents and folders would require less interaction and requests to a BSCW server if features for archiving and extraction of document archives were available.

These requirements resulted from our use of the Web for collaborative information sharing with BSCW, and in particular its use by one organisation. We believe however that many of the problems revealed are more generally valid: users suffer problems of bandwidth (in our organisation a common 'coping strategy' is to switch off image transfer at peak times); they need to download multiple documents; content providers often make information available in different formats. We now discuss a solution to these problems, first relating this to BSCW, but then demonstrating its applicability to the Web more generally.

A Content Transformation Assistant

The transformation assistant is a server-side component which provides an interface to conversion, encoding and clicking on the bag icon below the workspace listing [Fig. 1].

BSCW in Use: Requirements for Transforming Content

BSCW as described above provides no features for transforming content; successful information sharing requires use of document formats that collaborators have the necessary software to handle. We have evaluated this and other aspects of BSCW in the context of CoopWWW; a project in which BSCW forms the kernel of a more comprehensive Web-based groupware environment [Appelt 96]. CoopWWW involves cycles of development, deployment and evaluation with the project's user groups. We consider here findings from the first evaluation phase which focused on the use of the BSCW system by one user group, an organisation which provides scientific consultancy services. The organisation is distributed world-wide with offices in the UK, USA, Spain and Sweden.

Data was collected by way of questionnaires and a user workshop where problems were discussed with developers. One of the first findings was that users were having problems with data formats. Within the organisation PC, Macintosh and Unix platforms were in use and documents were being uploaded in application- and platform-specific formats which other users did not have applications (or correct versions) to handle.

To address this the organisation purchased a 'plug-in' utility for Netscape that interpreted different data formats and displayed the results in the Web browser. This required the administrators to install and update the software on all users' machines, and only allowed *viewing* documents in a workspace; *modification* still required the creator application, inhibiting collaborative document production. The organisation was then forced to establish conventions regarding 'permitted' data formats when there was a need for joint document production. These conventions also outlawed use of compression and archiving due to the lack of cross-platform availability of suitable utilities.

These conventions revealed a further set of problems, previously hidden when users could archive and compress documents before uploading. Bandwidth problems, first from the organisation's London office to a BSCW server at GMD then between satellite offices following installation of a server in London, showed that archiving and compression allowed optimal use of the channel to the server. Archiving required only one upload request to the server, while compression reduced transmission time considerably. The establishment of conventions necessary to achieve interoperability thus required more cumbersome and time-consuming upload and download procedures.

Analysis of these problems has identified the need for on-demand content transformation services:

- *Format conversion*: Conversion of workspace documents prior to download is required for users to view/edit documents for which users lack suitable application or conversion software. This would remove some of the need to install and maintain conversion and viewing software on all users' machines. Furthermore, users would have immediate access to new convertors following installation on the server.
- *Encoding/decoding*: With low bandwidth it is more efficient for users to upload [download] compressed data with the server [client] handling decompression. Conversely, it should be possible to decompress a compressed workspace document prior to download if that compression format cannot be handled by the user's machine.
- *Archiving/extraction*: Upload [download] of multiple documents and folders would require less interaction and requests to a BSCW server if features for archiving and extraction of document archives were available.

These requirements resulted from our use of the Web for collaborative information sharing with BSCW, and in particular its use by one organisation. We believe however that many of the problems revealed are more generally valid: users suffer problems of bandwidth (in our organisation a common 'coping strategy' is to switch off image transfer at peak times); they need to download multiple documents; content providers often make information available in different formats. We now discuss a solution to these problems, first relating this to BSCW, but then demonstrating its applicability to the Web more generally.

A Content Transformation Assistant

The transformation assistant is a server-side component which provides an interface to conversion, encoding and archiving utilities. The assistant does not implement specific utilities itself, but rather provides an architecture for adding utilities which the assistant can invoke to perform specific transformations on-demand.

Components of the Assistant

The assistant stores details of available conversion, encoding and archiving tools in a database. It provides three services: *consultation*, for information on possible transformations, *invocation*, for requesting transformations, and *administration*, for updating the database (e.g. when a new conversion tool is installed). Clients can use the consultation interface to ask:

- What conversions, encodings and archiving operations are possible for documents of format *F*?
- Is it possible to transform a document of format *F1* into format *F2*?
- (In each of the above cases) what is the 'quality' associated with the transformations? Are features lost?

The assistant uses the property database to answer these questions. MIME-types identify the source and target formats. To request a transformation, the client uses the invocation interface to specify the location of the source document(s) and source and target formats to identify the transformation required. The assistant then performs the transformation returning the location of the new documents (or directories, in the case of extracting an archive).

To add new transformation tools the administrator provides details through the administration interface. To add a new convertor which translates LaTeX documents to HTML for example, the administrator would specify:

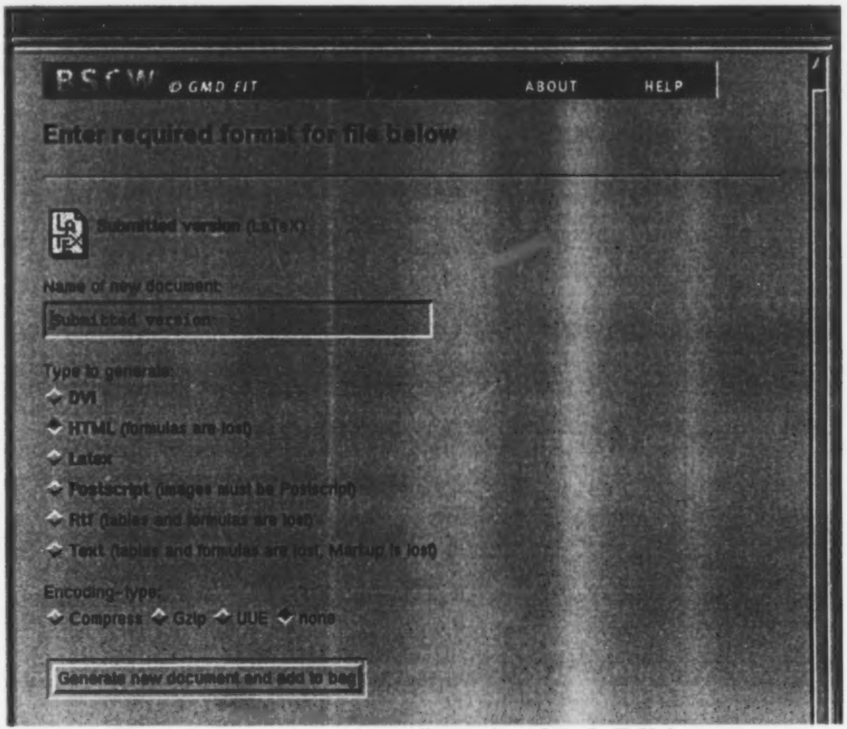
- The MIME-type of the source format ('application/latex')
- The MIME-type of the target format ('text/html')
- The method of invoking the convertor (e.g. '/usr/local/bin/latex2html <src> <tgt>')
- Information on the conversion 'quality' (e.g. "formulas are lost")

The assistant uses the database to build a directed graph representing transformation possibilities. The nodes are MIME-types of different data formats, the links represent particular transformation methods annotated with the quality information describing the characteristics of the transformation. More than one method of performing the same transformation can be given, perhaps invoking the same tool with different parameters. This makes sense if the conversions have different characteristics, reflected in the 'quality' information provided to give an idea of the limitations or degradation resulting from using a convertor. If multiple direct transformations are possible between formats, then multiple directed links will also exist between corresponding nodes in the graph. Both consultation and invocation then involve simple graph traversal, the former to see which paths exist and gather associated quality information, the latter to invoke the associated transformations.

Adding Content Transformation Services to BSCW

The transformation assistant has been implemented in Python and integrated with BSCW. The interface has been augmented with a 'convert' operation for each document object and image file, and an 'archive' operation which can be applied to multiple object selections. These operations invoke requests to the BSCW server which are forwarded to the transformation assistant's consultation interface. BSCW uses the information returned to generate a page of possible transformations for the selected object(s) plus associated quality information. The user can then invoke a transformation operation and afterwards download the results and/or store them in a workspace.

Requesting to 'convert' the LaTeX document "Submitted version" returns the form shown in [Fig. 2]. BSCW looks up the MIME type of the document in its own database and consults the assistant on possible conversions for "application/latex". The server also asks for encoding possibilities which could be applied after or instead of any conversion. BSCW sends the selected transformation to the assistant's invocation interface which performs it, returning the path of the resulting file. Archiving is similar. BSCW again consults the assistant to return a list of archiving options plus applicable encodings. If folders are selected their contents are also added. When compiling the archive, BSCW checks the access rights for the user with respect to objects within a folder hierarchy, only adding objects the user is permitted to read. Objects such as URLs and Articles are transformed to plain text files before archiving. For each invocation of the transformation assistant, BSCW adds the resulting file as a document object to the user's Bag, from which the user can download it and/or move it to a workspace to make it available for workspace members [Fig. 3].



Enter required format for file below

Submitted version (LaTeX)

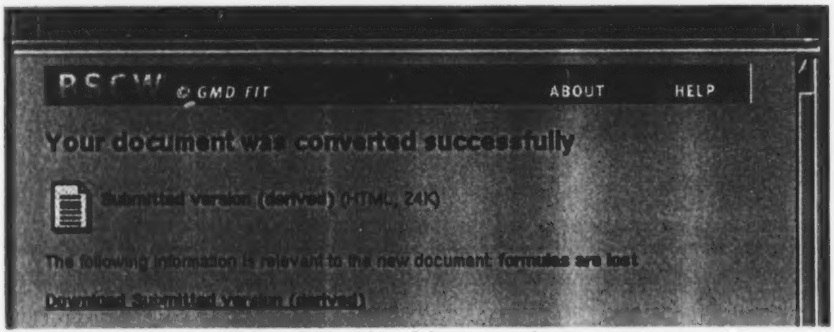
Name of new document:
Submitted version

Type to generate:
○ DVI
● HTML (formulas are lost)
○ Latex
○ Postscript (images must be Postscript)
○ Rtf (tables and formulas are lost)
○ Text (tables and formulas are lost, Markup is lost)

Encoding-type:
○ Compress ○ Gzip ○ UUE ● none

Figure 2. Conversion and encoding options for a LaTeX document

This implementation addresses some of the requirements identified in our evaluation of BSCW. An enhanced version offering further possibilities for content transformation, e.g. extraction of uploaded archives, is currently being implemented, along with a Web-based tool for administering the transformation assistant.



Your document was converted successfully

Submitted version (derived) (HTML, 24K)

The following information is relevant to the new document: formulas are lost

Download Submitted version (derived)

Figure 3. Results of the conversion process

Conclusions: Content transformation and Content-negotiation

The Web community recognises the general need to provide content in different forms. The term 'content negotiation' [Behlendorf 96] refers to methods by which browsers and servers can select from 'variants' of the same information 'resource'. Current work in this area is defining standards for specifying and selecting from multiple

BEST COPY AVAILABLE

variants with different properties based on user preferences. We do not have space here to discuss the details of the proposed scheme (see [Holtman & Mutz 97] for details), and thus limit ourselves to high-level comparisons.

The emphasis of the proposed method is on 'transparent' selection from a list of variants, in contrast to BSCW which supports only explicit selection. This is shown by our use of a 'quality string' to describe possible degradations [Fig. 2] rather than a quality 'factor' as used by the content-negotiation algorithm to select the 'best' variant for a user's stated preferences. Often the 'quality' of a variant cannot be divorced from its intended use; a document variant which has lost all formatting may be useless for printing and distribution but adequate if what is required is to cut and paste a paragraph of text. With the scheme proposed in [Holtman & Mutz 97] it is possible to use the 'variant description' for such information, and we believe that browsers and servers should support this.

A utility of our approach is to remove the need to provide multiple variants of each resource. For example, in the current Apache server implementation of content negotiation, the server looks for variants of a resource by consulting a configuration file or looking for files in the same directory as a requested document with the same name but different suffix. In both these cases the provider must manually produce each variant of the resource.

We have therefore extended the transformation assistant to generate variant information in the form described in [Holtman & Mutz 97]. A quality factor is also stored in the database and quality descriptions are returned in the 'variant description' field. With this approach the server consults the assistant for information on variants which can be dynamically created and returns this information to the browser as a variant list. This means that, with appropriate transformation tools, new documents added to a Web site are automatically available in different variants and making all existing documents available in a new format requires only adding a suitable conversion tool.

Both manual and automatic transformation approaches are compatible with the proposed content-negotiation scheme, and a combination might be useful; for example, if all that is required is a rough outline of a document in a different language, invocation of a language translation utility might be sufficient. More accurate translation will require the skill of a human translator and the creation of multiple, static document variants.

Acknowledgments

The work reported in this paper was funded as part of the CoopWWW project by the European Union under the Telematics Applications Programme, grant TE 2003. We wish to thank Matthias Jarke and his group from the Technical University of Aachen who performed the evaluation studies, as well as Sanjeev Shah and Gerd Woetzel who provided valuable input to this paper.

References

[Appelt 96] Appelt, W. (1996). CoopWWW - Interoperable tools for cooperation support using the WWW, *5th ERCIM/W4G Workshop: CSCW and the Web*, Sankt Augustin, Germany, April 1996, 95-99. Available from GMD as Working Paper #984 or on-line at: http://orgwis.gmd.de/projects/W4G/proceedings/coopwww.html

[Behlendorf 97] Behlendorf, B. (1997). *What is Content Negotiation? Towards an extensible framework for an ecology of data types*, available from: http://www.organic.com/staff/brian/cn/

[Bentley et al. 97] Bentley, R., Appelt, W., Busbach, U., Hinrichs, E., Kerr, D., Sikkel, K., Trevor, J. & Woetzel, G. (1997). Basic Support for Cooperative Work on the World Wide Web, *International Journal of Human-Computer Interaction*, 46(6), Academic Press, Cambridge, June 1997, 827-846.

[Holtman & Mutz 97] Holtman, K. & Mutz, A. (1997). *Transparent Content Negotiation in HTTP*, Internet Draft, available from: ftp://ftp.ietf.org/internet-drafts/draft-ietf-http-negotiation-03.txt

Website Authoring using Dynamic Hypertext

Rick Bodner, Mark Chignell, and Jim Tam
Mechanical and Industrial Engineering
4 Taddle Creek Road
University of Toronto
Toronto, Ontario, M5S 1A4
Canada
E-mail: [rbodner, chignel, tam]@ie.utoronto.ca

Abstract: Traditional (static) hypertext requires an author to hardcode links between documents. Users see a "one size fits all" organization which may not meet their searching needs. Dynamic hypertext creates hypertext links "on the fly". These links customize a Web site uniquely for each user. An experiment was conducted comparing dynamic and static forms of hypertext in a question answering task. Overall performance, in terms of both task time and accuracy, was better for the dynamic hypertext interface. In addition, novices appeared to benefit more from using the dynamic hypertext than experts. As well as benefiting novice users, dynamic hypertext is easier to author, since the links are not created a priori. There is no longer a need to check the integrity of links and/or update them, and new documents may simply be added to the Web site without needing to integrate (i.e., link) them with other documents within the Web site.

Introduction

Web site authoring is currently a time consuming process with most Web sites being largely crafted by hand. Originally, sites were constructed by hardcoding HTML in text editors. More recently, there has been a move towards building Web pages with WYSIWYG Web authoring tools. However, major effort is still required to build individual pages, and to create and maintain the hypertext links (which will be referred to as hyperlinks or simply links) between various pages within a site. Furthermore, for a large site it may be impossible for authors/Web designers to anticipate all the information needs that visitors may have, let alone anticipate all those needs with appropriate links and navigation options.

Search engines and meta-indexes (such as Yahoo) have become major parts of Web navigation. This is largely due to the defects of simple point and click browsing using links. Large amounts of searching are necessary because there is usually no clear path that a user can take to get from a current location to a desired location, if the URL (Uniform Resource Locators) for the target location or concept is not already known. Furthermore, since authoring of hypertext links is highly dependent on the whims and biases of the author, and since paths of interest may cross boundaries between the work of one author and another (each with their own interests and concerns), navigation solely by links on the Web tends to be a matter of serendipity more than planning and successful completion of navigational plans.

Thus it is not surprising that a common information exploration strategy on the Web is to begin with a search, visit search results until a relevant site is found, and then navigate through that site with hyperlinks, returning to the search engine when it is time to look for further information. Alternatively, if the site is central to the topic and well-connected to related material, a person may use the site as a launching point to create a "spiky" navigation pattern, as described by [Campagnini & Ehrlich, 1989], [Parunak, 1989], and others.

Although the dominant structural feature of the Web was the hyperlink, in practice searching based on queries of one form or another has become just as important as the hyperlink in exploring information on the Web. Current implementations of search engines create two distinct modes in Web navigation, i.e., submitting a query to a search engine, and hyperlinking. However, there are good theoretical reasons to believe that combining browsing and querying more effectively [see Waterworth & Chignell, 1991] may lead to improved Web navigation.

In this paper we describe the DynaWeb system for constructing Web pages automatically and dynamically from large text databases, based on user inputs. We will also review experimental results that have been obtained using the system. The DynaWeb system has been developed at the University of Toronto on the basis of a dynamic linking method originally developed by [Golovchinsky, 1996] in the VOIR electronic newspaper interface prototype. It is suggested that dynamic hypertext construction on the Web (using the DynaWeb approach) will not only reduce Web authoring effort, but may also result in more useful Web pages in many circumstances.

Overview of Dynamic Hypertext and the DynaWeb System

Currently, hyperlinks found on Web pages are static. That is to say, the URL (e.g., another Web page) specified by the link only changes when the author of the page wishes it to change. This gives excessive control to the author over what people experience while browsing the Web site. Each person who visits the Web site is presented with the same set of links and the manner in which these links are organized may inhibit the person from obtaining the sought after information.

In a dynamic version of hypertext, no static links exist in a Web page and the links are created "on the fly" These dynamic links are created based on the content of the Web site and the interests of the user. Depending on these criteria, certain words or phrases are made into links. These links could lead to other Web pages in the Web site or can be used to further a user's search. Dynamic hypertext can be used to present conventional text databases as a set of interconnected Web pages, and to merge them with other information on the Web (e.g., conventional static hypertext pages). For a further discussion of dynamic hypertext [see Golovchinsky, 1997a and Golovchinsky, 1997b].

The basic concept of dynamic hypertext was implemented in the DynaWeb system [Fig. 1]. The DynaWeb system is intended to be used to present large textual databases as Web pages. The system interacts with the INQUERY search engine, developed at the University of Massachusetts [Caltin *et al.*, 1992], to retrieve relevant document(s) in response to a query. The titles of the most relevant documents (e.g., the top ten) are presented [Fig. 2]. When the person selects a document, the document is presented to the person with hyperlinks which were created "on the fly" [Fig. 3]. The selection of words used as links is based on previous links (i.e., queries) the person selected. The system also keeps track of the previous queries.

The core of the DynaWeb system consists of a set of CGI programs implemented in C which communicates with the INQUERY search engine and composes (i.e., creates the links for) Web pages and presents them to the user. These programs also receive the selected link information from the user. A running query is further refined using the surrounding text around the link and adding it to the text from two previously followed links. The DynaWeb system requires this information to try to build a profile of the user's information need. In this way the system is able to tailor the links to match the needs of the person. The words selected as links then point to a menu of likely documents that can be easily scanned. Presenting the information in this way means that the searching task is no longer recall oriented, but recognition oriented, using a point and click (hypertext links) browsing interface.

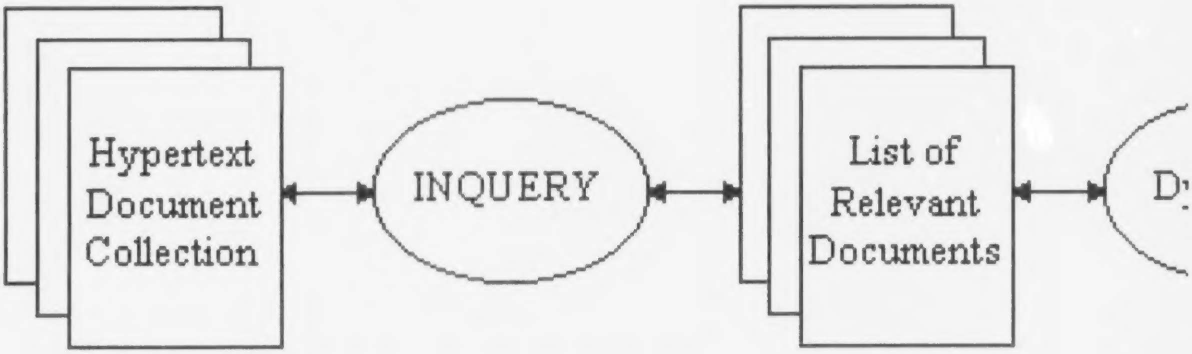
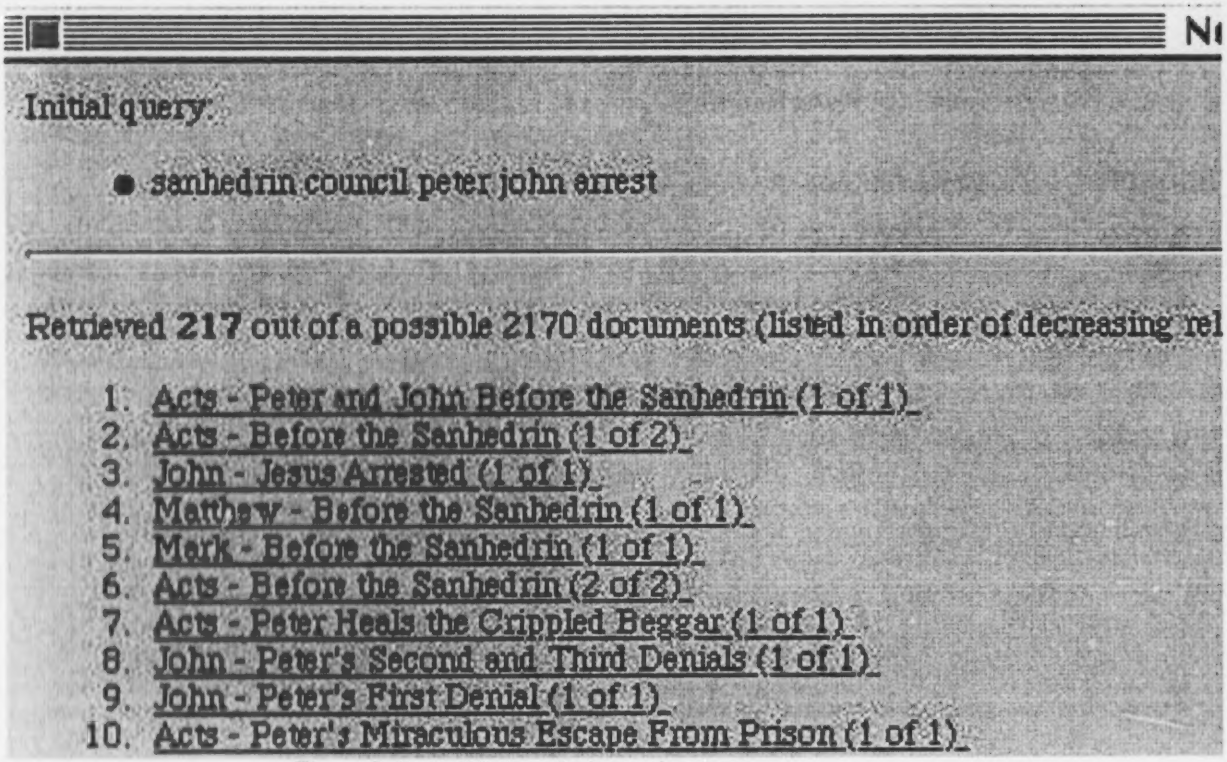


Figure 1: Structure of the DynaWeb system.



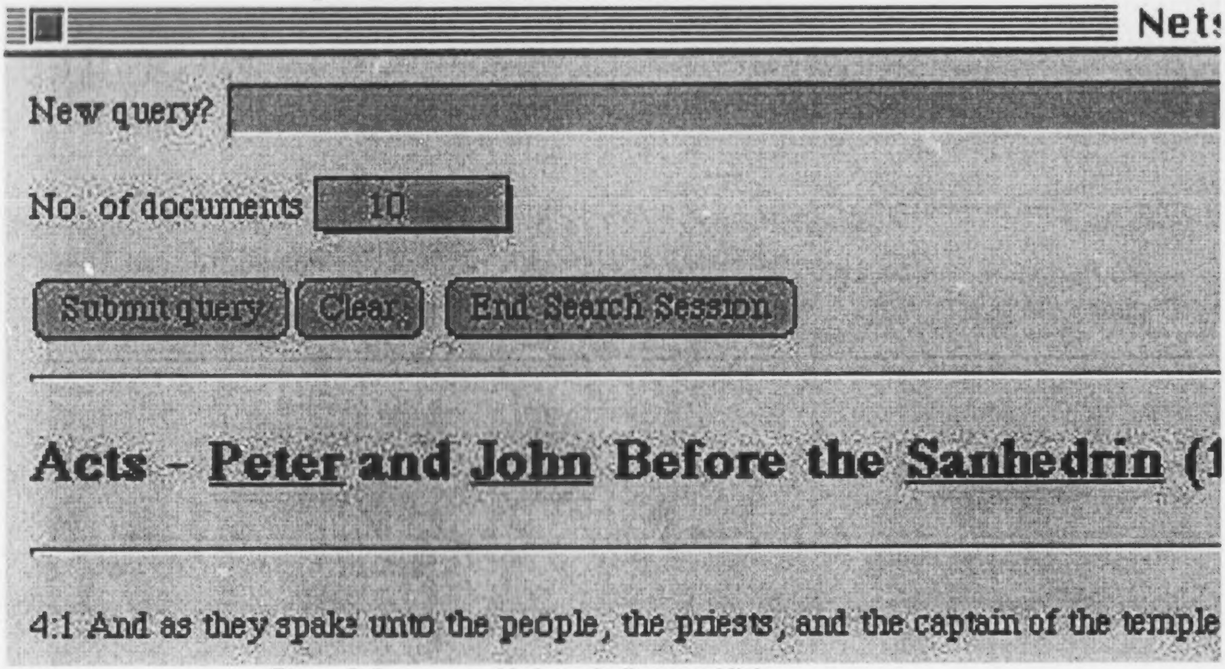
Initial query:

- sanhedrin council peter john arrest

Retrieved **217** out of a possible 2170 documents (listed in order of decreasing rel

1. Acts - Peter and John Before the Sanhedrin (1 of 1)
2. Acts - Before the Sanhedrin (1 of 2)
3. John - Jesus Arrested (1 of 1)
4. Matthew - Before the Sanhedrin (1 of 1)
5. Mark - Before the Sanhedrin (1 of 1)
6. Acts - Before the Sanhedrin (2 of 2)
7. Acts - Peter Heals the Crippled Beggar (1 of 1)
8. John - Peter's Second and Third Denials (1 of 1)
9. John - Peter's First Denial (1 of 1)
10. Acts - Peter's Miraculous Escape From Prison (1 of 1)

Figure 2: Ranked listing of relevant documents.



No. of documents 10

Acts - Peter and John Before the Sanhedrin (1

4:1 And as they spake unto the people, the priests, and the captain of the temple

Figure 3: Document with dynamically created links.

User Study

BEST COPY AVAILABLE

This section describes one experiment conducted in our laboratory that compared a DynaWeb implementation with an equivalent static hypertext [Tam, 1997]. The experiment studied user performance (in terms of task time and accuracy) while performing a question answering task. In the static linking condition, links were implemented in the research prototype in accordance with the cross-referencing provided by the Thompson's Bible Chain References [Thompson, 1934]. In contrast, the dynamic linking condition presented links created "on the fly" using DynaWeb's heuristic search-based algorithm. Within each interface condition (static or dynamic), subjects were required to answer eight questions, of which four were selected to be factual and four were selected to be analytical. (Factual questions were those whose answers could be found within a single Bible passage. Analytical questions were those with answers that needed to be deduced from analysis of multiple Bible passages).

Twenty people participated in the experiment. Participants were randomly assigned into two groups (one for each interface order, with five novices and five experts in each group). The novice subjects were undergraduate or graduate students (from a wide variety of disciplines) from the University of Toronto. The experts were mostly students or recent graduates from the Ontario Bible College, the Ontario Theological Seminary or the Wycliffe Seminary at the University of Toronto. Each subject received an instruction sheet and training on how to use the interface to perform the question-answering tasks. This was later repeated when the subject switched to using the second interface (after completing the first eight questions). The subjects performed eight question-answering tasks using the first interface (either static or dynamic, depending on which group they were in). They then took a short break before using the second interface to perform another set of eight question-answering tasks.

The data were analyzed using analysis of variance (ANOVA). For overall task time there was a significant interaction of interface and expertise (F[1,18]=15.24, p=.001) as shown [see Fig. 4]. The dynamic interface reduced the task time for novices substantially, but had little or no effect on the task times of experts.

For analyzing the accuracy data, the results for the questions were pooled across each of the four combinations of interface and question types. The accuracy scores (out of a possible four) were obtained by summing separately over the four factual questions and the four analytical questions answered in each of the interface conditions (dynamic vs. static). For the accuracy scores, there was a significant three way interaction between interface, question type and expertise (F[1,18]=29.72, p<.001).

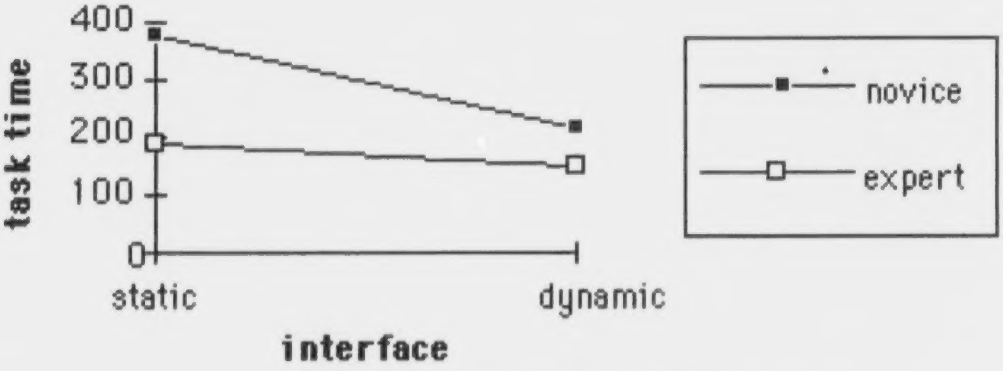


Figure 4: Task time (interface by expertise).

For the Dynamic interface there was little difference between novices and experts on accuracy, whereas for the static interface novices were less accurate than experts (F[1,18]=56.0, p<.001), particularly for the factual questions (for the interaction of expertise and question type, F[1,18]=44.81, p<.001).

These results indicate that dynamic linking as implemented in DynaWeb can improve question answering performance, particularly for people who are not very familiar with the search domain. In the study that we carried out, the use of dynamic linking tended to improve the performance of novices relative to that of experts, both in terms of task time and question answering accuracy.

Web site Authoring and Maintenance

As mentioned above, much time is spent creating and maintaining static hypertext links. A Web site (a collections of hypertext documents) is dynamic over time. That is to say, the links between the documents are continuously updated and/or new ones are added. This can be due to many reasons, such as new documents being added, or the accommodation of changing interests. For example, a corporate Web site will need to be updated when a new product is introduced. The addition of this document may have a ripple effect throughout the company as different departments add or change links to reference this new document. These ripple effects can also extend outside of the company. A dealer for the company may also have to update their Web site to reference the new document. This simple example can be extrapolated to the World Wide Web, where millions of people are creating and updating Web sites, and where these ripple effects can become quite problematic.

Dynamic hypertext was not intentionally designed to address this ripple effect problem, but can be used as a means to control it. Since links are created "on the fly" in a dynamic hypertext system (e.g., the DynaWeb system), there is no need for hardcoding links. This means that the documents used to make up a Web site do not contain any hyperlink tags. The only tags that author need to concern himself/herself with are the ones related to formatting the text. Generally, the formatting of a document is done once when the document is created. The emergence of style sheets as a factor in HTML authoring will further simplify the task of document formatters of a large Web site. Once created the document is simply added to the Web site and the dynamic hypertext system takes care of the links between the documents. This can potentially reduce the amount of time a person spends authoring a Web site. No longer does the author need to be concerned with linking new documents to existing ones or with checking on the integrity of existing links.

In addition, Web sites have to be maintained in such a way that they are well structured and connected. Marsh (1997) found that the structure of a Web site has a large impact on how users navigate in the site, the memorability of the site, and its perceived size. In addition, she found that hypertexts with nonhierarchical structure appeared to be smaller than strictly hierarchical sites with a similar number of nodes. Subjects tended to visit more nodes within strictly hierarchical sites than hierarchies that had cross links between branches. Since Web sites allow any node to be linked to any other, the complexity of the resulting network can increase arbitrarily. This explains why most large Web sites tend to be hierarchical with relatively few links between branches. Consequently, the maintenance efforts in terms of time and costs would be tremendous. With dynamic linking, effort required to ensure that the site is well structured with well connected links can be reduced to a bare minimum.

Conclusions

There is considerable skepticism about the effectiveness of static hypertext linking methods (e.g., [McKnight *et al.*, 1989]. Dynamic linking promises to address these limitation issues by creating hypertexts that are responsive to users' needs and interests. We expect that dynamic links will be particularly effective in:

- reducing disorientation
- lowering cognitive overload
- enabling flexible access
- reducing authoring effort
- allowing users to express themselves more
- improving poor structure in hypertext
- making a wider range of the hypertext "visible" through globalization.

With dynamic linking, documents can be organized based on the information need of the searcher. The Web site author need not impose his/her whims or biases on the searcher (through the organization they impose in creating static links). Instead, the organization of the presented information is dynamic, and interactively defined by the searcher. In contrast, static Web sites generally rely on graphics and forms to provide this level of interactivity.

We have gained considerable experience in the development of dynamic hypertexts in our laboratory. In general, we have found that the only formatting needed to convert typical text documents into a dynamic hypertext is to add HTML tags for the title, and for identifying new paragraphs. This task can usually be easily automated given the structured nature of most documents. Once the documents are collected into a database, the DynaWeb system takes care of creating links (building the organization) between the documents. Much more effort typically goes into creating an equivalent static hypertext. For instance in the static hypertext version of the Bible used in Tam's study [Tam, 1997], figuring out a method to create the links between the text documents required a great deal of work. This effort would have been much greater still if we did not have the static links already defined by the Thompson chain references. Furthermore, if another document collection is added to the static corpus, much more effort is required to integrate (link) new and old documents. In contrast, with dynamic Web corpus there is no need to worry about the addition of new documents since no links have to be updated.

Dynamism provides flexible and user-customized access to information. With links generated "on the fly" users do not have to follow predetermined (and fixed) links. There are more access or entry points into the hypertext, depending on the context of the exploration tasks. Dynamic links also have the desirable property of requiring no authoring effort other than indexing the text database made up of all the hypertext nodes.

The research reported in this paper used a particular dynamic linking strategy originally developed by [Golovchinsky, 1996]. Tam's study [Tam, 1997] showed how dynamic linking improves performance in a question answering task. Earlier, [Golovchinsky, 1996] showed that dynamic linking can improve performance in an information retrieval task (where the goal is to obtain a set of highly relevant documents in response to a particular search topic). Tam's study looked at domain expertise (Bible knowledge) and showed how dynamic linking helps domain novices in particular. Golovchinsky's research showed that dynamic linking also benefited search novices (i.e., people who had little if any experience in searching online databases) much more than search experts (e.g., professional librarians and search intermediaries). Thus dynamic linking can "level the playing field" by helping novices to explore information more effectively. It also allows any collection of text documents to be repurposed into dynamic hypertext with little if any authoring effort.

Acknowledgments

We gratefully acknowledge support and funding for our research laboratory from the following sources: the National Science and Research Council of Canada (NSERC) and the Information Technology Research Centre of Ontario (ITRC).

References

[Caltan *et al.*, 1992] Caltan, J. P., Croft, W. B., and Harding, S. M. (1992). The INQUERY retrieval system. In *Proceedings of the 3rd International Conference on Database and Expert Systems Applications*. 78-82.

[Campagnini & Ehrlich, 1989] Campagnini, F. R., and Ehrlich, K. (1989). Information retrieval using a hypertext-based help system. *ACM Transactions on Office Information Systems*, 7 (3), 271-291.

[Golovchinsky, 1996] Golovchinsky, G. (1996). *From information retrieval to hypertext and back again: The role of interaction in the information exploration interface*. Ph.D. Dissertation, Department of Mechanical and Industrial Engineering, University of Toronto.

[Golovchinsky, 1997a] Golovchinsky, G. (1997). Queries? Links? Is there a difference? In *Proceedings of CHI'97*, Atlanta, GA. ACM Press. 407-414.

[Golovchinsky, 1997b] Golovchinsky, G. (1997). What the query told the link: the integration of hypertext and information retrieval. In *Proceedings of Hypertext'97*, Southampton, UK. ACM Press.

[Marsh, 1997] Marsh, A. (1997). *Structure and memoriability of web sites*. Unpublished M.Sc. dissertation draft, Department of Mechanical and Industrial Engineering, University of Toronto.

[McKnight *et al.*, 1989] McKnight, C., Dillon, A., and Richardson, J. (1989). Problems in hyperland. *Hypermedia* 1 (2), 167-178.

[Parunak, 1989] Parunak, H. V. D. (1989). Don't link me in. In *Proceedings of Hypertext'89*. Pittsburgh, P.A. ACM Press. 43-50.

[Tam, 1997] Tam, J. (1997). *Design and evaluation of web-based dynamic hypertext*. Unpublished Ph.D dissertation draft, Department of Mechanical and Industrial Engineering, University of Toronto.

[Thompson, 1934] Thompson, (1934) *Bible Chain References*, Cross Country Software, Guelph, Ontario, Canada.

[Waterworth & Chignell, 1991] Waterworth, J. A., and Chignell, M. H. (1991). A model for information exploration. *Hypermedia* 3 (1), 35-58.

How to Make Money with Web Access

Manfred Bogen <Manfred.Bogen@gmd.de>
Guido Hansen <Guido.Hansen@gmd.de>
Michael Lenz <Michael.Lenz@gmd.de>

GMD - German National Research Center for Information Technology
D-53754 Sankt Augustin

Abstract: More and more businesses are done through the Internet. One can shop in virtual malls worldwide and one can purchase almost everything through the Internet and the Web. Internet Service Providers all over the world make money with Internet access t everybody. Paid advertisement on Web pages is normal nowadays. The idea to make money with Web access is rather new.

Late 1995, GMD scientists had the idea to develop W3Gate, a mail gateway to fill the gap between email and WWW. Under the influence of the experience made and the demand of different organizations to have the software, it was finally decided to start a commercially oriented project based on W3Gate. The goal was to get additional funding for maintenance and further development. This paper describes an approach to turn an academic project into a commercial one together with the additional requirements imposed by the new paradigm.

Keywords

Multimedia, multimedia network interface, communication protocols, distributed multimedia systems, electronic mail, WWW, hyper link documents, hypermedia, electronic commerce, security

Introduction

W3Gate is a gateway between email and WWW that enables email users to retrieve multimedia objects out of the Web. It is independent of the email protocol used, be it Internet or OSI mail. W3Gate is a very popular and it seems to be superior to other implementations of email-based Web access today [WebMail 95, Secret 94].

W3Gate was originally designed for people with a bad Internet connectivity, especially in less developed or developing countries and for people that don't have the ability to retrieve documents directly via http. Several scenarios were expected: they are not connected directly to the Internet but to a provider that offers only email access or they are connected to an Intranet that is guarded towards the Internet by a firewall. For security reasons only email but no direct Web access is allowed. From our usage statistics we quickly learned that even users with a rather good Internet connectivity can benefit from our service especially when trying to retrieve huge documents during normal business hours [Fig. 1]. In the meantime, we receive urgent requests to get the software for free or to learn about the distribution policy and financial conditions.

This paper structures as follows: after a short introduction of W3Gate's functionality, [Filling the Gap] describes its usage. [Quality of Service] presents the different quality of service requirements we met. [Electronic Commerce] introduces our present marketing strategy. The [Conclusion] tells that we did not make any dime with W3Gate up to now. However, we are confident to have better results until the time of the WebNet'97 conference in Toronto.

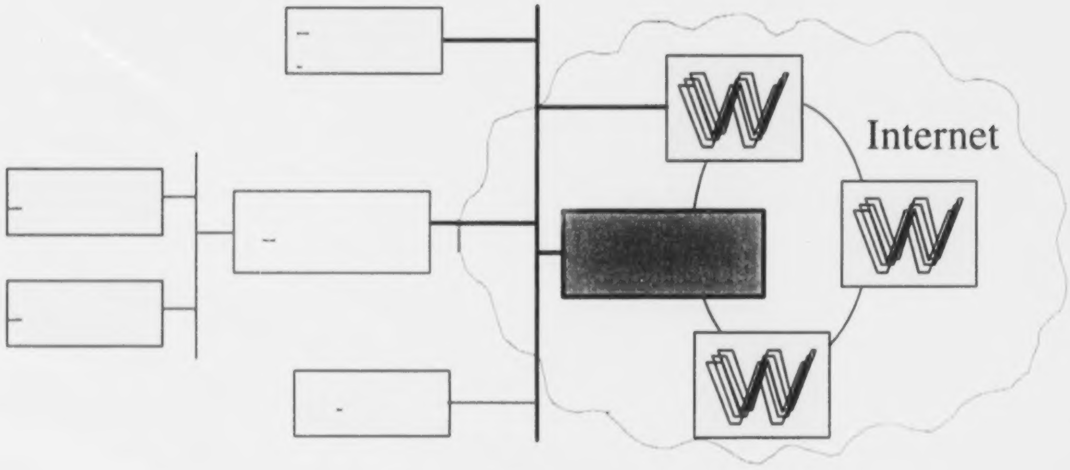


Figure 1: W3Gate Overview

W3Gate's Functionality

An email sent to *w3mail@gmd.de* can request WWW and FTP documents [Fig. 1]. Following [Berners-Lee, Masinter, McCahill 94] specifying the respective userid and password within their URLs can even retrieve protected documents. An email may contain up to 10 get requests. The subject field of the mail is currently unused. The commands are case-in-sensitive. Without any option, the *get* command returns the document denoted by *URL* unchanged to the requester.

help	request a help document
get [-t | -u | -a [-c columns]] [-ps] [-z] [-uu] [-s size] [-img] [-l] URL	
	request document denoted by URL
-t	strip all tags
-u	preserve links to other documents as relative URLs if possible
-a	preserve links to other documents as absolute URLs
-c *columns*	wrap lines after *columns* columns
-ps	convert ASCII text into postscript format
-z	compress document
-uu	uuencode before mailing
-s size	set size of document in email to size [Kbytes]
-img	get all inline-images
-l	get all documents from links

Table 1: W3Gate Commands

The options *-t*, *-u*, and *-a* are mutually exclusive. If one of these options is present, the requested document is formatted according to the HTML tags [Ragget 96] included, if any. If one of the *-u* or *-a* options is specified, all URLs to linked documents are preserved in the text either as relative or as absolute URLs.

The *-c* option is only allowed in conjunction with one of the options *-t*, *-u*, or *-a*. If specified, the document is formatted with the given number of columns. If the value remains under 40 or exceeds 255 it is set to the respective limit. The number of columns defaults to 80.

The *-ps* option causes any ASCII document to be converted into a postscript document. The document is displayed in portrait mode. Large documents can be compressed (*-z*) and uuencoded (*-uu*) before mailing.

Users can specify the maximum message size in Kbytes that their electronic mail system can handle using the *-s* option. By specifying *-l* W3Gate not only fetches the document denoted by the URL itself but also all other documents referenced by hyper links within this document. So it is possible to browse through a series of hypertext documents off-line. Analogously the *-img* option tries to fetch all included images in a document. Hereby users get an optical impression of the Web site.

Filling the Gap

A successful business model to sell products or to offer services especially through the Internet consists out of the following components:

- A high-quality product or service with a strategic position in the market ('where is the gap?'),
- Satisfied customers and clients who trust in the product or the service,
- Reliable, enhanced, and sophisticated usage or purchase statistics,
- Eventually a secure appropriate payment method,
- A reliable financing model, and
- A powerful marketing strategy.

Since W3Gate started its work in May 1995, the number of transmitted files grew from a few to an average of 17.000 files a day. Simultaneously the traffic grew up to 9 GB per month [Fig. 2]. By now, W3Gate is a well-accepted service that is used by many users on a regular basis.

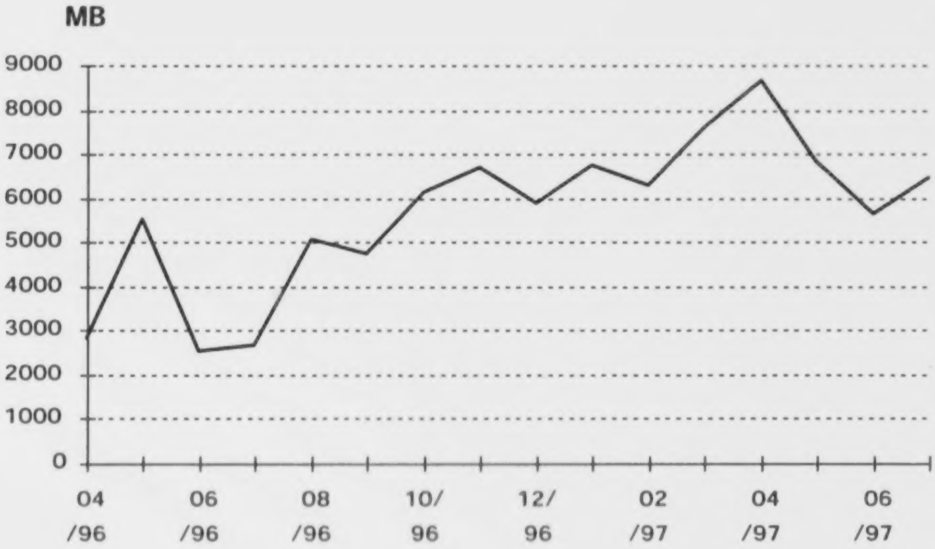


Figure 2: W3Gate traffic

Our enhanced statistics show that W3Gate's users come from more than 80 countries. They are from universities and schools, from profit organizations and companies like IBM, Sun, SONY and Nike, and also from organizations like the United Nations and the World Bank. Last month, 19% of the requests were from Germany and 18% from the US. Not only plain HTML-files, but also binary files had to be processed by W3Gate.

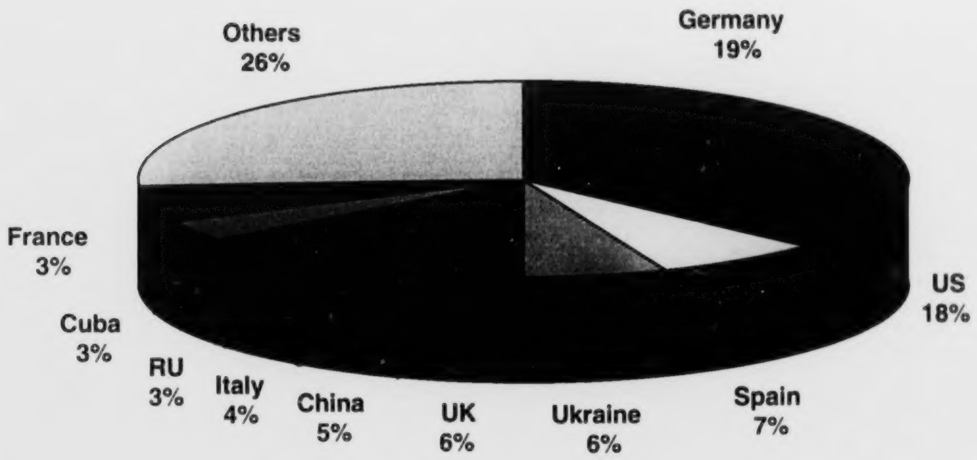


Figure 3: Countries using W3Gate

This means that we found a gap in the market, that we have regular customers who really rely on our development, and that we know almost everything about the use and also abuse of W3Gate. However, how good is W3Gate? Is it of high quality?

Quality of Service

On its way to a regular, well-accepted service, W3Gate had to meet different quality of service (QoS) expectations.

Availability

In the early days, W3Gate was taken out of service for a couple of times to fix actual problems and to clean up our machines. All requests during this time were lost, sometimes without informing our customers. As users simply expect a service 24 hours on 365 days of a year, we had to change this in the latest version (2.0).

Reliability

Second, a high reliability is very important for a commercial service. For each received request a corresponding reply has to be sent, be it the requested document, an error report, or a help file explaining the commands and their usage. Even in case of system crashes, no request received should be lost and no reply should be sent multiple times.

In order to achieve these goals we changed W3Gate's design. Now all requests and their current state are persistently stored in the file system. Hence, after a restart, W3Gate can reconstruct the system status out of these files. On unavailability of documents, requests are repeated a several times in irregular intervals.

Usability

Third, user friendliness and ease of use play an important role for the service acceptance. Incorrect commands received generate a message that cites the wrong commands and explains the nature of the error. Mail signatures that are preceded by an '-- ' (minus+minus+blank) [Spencer 94] in a separate line are ignored.

BEST COPY AVAILABLE

With the *-I* or *-img* option specified, W3Gate generates an index list at the beginning of the returned message with the URLs of the documents attached to the message. Using a Web browser like Netscape's Navigator or Microsoft's Internet Explorer, users can click on these links and their browser displays the document immediately.

Especially old mailing systems can only handle properly messages up to a specific maximum size, which often happens to be 100 K. Thus, W3Gate automatically splits larger documents on default into several messages of that size, but the actual partial message size can be specified by the -s option. From the *subject*-field users can see in which order the whole file has to be reassembled. They just have to strip off the header-lines from each message and join the parts together according to the sequence number.

User Support

Finally, we established a W3Gate administration email address (w3mail.admin@gmd.de). All comments related to the service as well as requests for help can be sent to this address. They will be handled immediately by our administrators.

Electronic Commerce

The components still missing in our business model were the secure payment method, a reliable finance model, and a powerful marketing strategy. To make progress in this, we needed to talk to marketing experts. We started our discussions with co-operation partners and potential resellers of the software, presented our early marketing ideas, and had to redefine the strategy finally.

The Early Plans

Basically, we wanted to work with identified users. In former versions of W3Gate, it was possible to request documents on behalf of someone else. On one side, this led to complaints by innocent users, on the other side you can only send an invoice to users if they are real customers. The best way to solve both problems is a closed system for registered users with a user registration using an email-based three-way-handshake protocol. While this approach could not give us a complete security against malicious users, it should help us to know our customers and to work with valid email addresses, not only for billing. Additionally, an integration of PGP [PGP 97] or PEM would make sense.

Finally, we wanted a cost-effective, convenient, and rapid solution for processing financial transactions on the Internet to pay W3Gate's invoice. Credit card numbers over the Internet seemed inappropriate. Instead we looked for an electronic payment system which should preferably be email based ,as this is the basis for W3Gate. We did not want a cumbersome administration for invoices of a few dollars only. Therefore we planned to work with monthly invoices.

The New Marketing Strategy

In our discussions with the marketing experts, we very soon found out that we definitely do not want to establish a big accounting and billing bureaucracy for W3Gate. We did not really expect people in developing countries to pay our monthly bills. By this, we also could neglect our ideas about identified users and digital signatures.

Nevertheless, we wanted to follow two commercial paths: sell the software to interested parties like mail-service providers and companies, and establish a service on behalf of a sponsoring third party, interested in W3Gate's feature to attach commercial information to the HTML documents originally requested. By this, we could forget our ideas about electronic money.

Additionally it seems appropriate to talk to firewall manufacturers who can use W3Gate as a proxy for their firewall software.

We finally found a co-operation partner with promising contacts willing to support us in following this commercialisation path. A test installation of the W3Gate software will be made there and a license agreement for the software sale and a co-operation agreement related to a sponsored third-party-service are in preparation. The distribution of tasks is rather straight forward: GMD will be responsible for the W3Gate technology, our co-operation partner for the marketing and advertisement.

Conclusion

Having done our homework related to quality of service and security issues as described in this paper, we found our users and a gap in the market. Our marketing strategy has changed and fortunately less of technical work has to be done in consequence. The operational experience of W3Gate during the last two years makes us trust in the software and confident to have more recent results at the end of 1997 on occasion of the WebNet'97 conference.

References

[Altmayer-Pizzorno, Bogen, Wermelskirchen 95] J. Altmayer-Pizzorno, M. Bogen & C. Wermelskirchen (1995). *E-Mail Quality as a Prerequisite for CSCW Applications*, 1995, Proceedings of JENC6, Tel Aviv.

[Berners-Lee, Masinter, McCahill 94] T. Berners-Lee, L. Masinter, M. McCahill (1994). *Uniform Resource Locators*, RFC1738.

[Bogen, Hansen, Lenz 96] M. Bogen, G. Hansen, M. Lenz (1996). *W3Gate - A Web Access for Outsiders*, 1996, Proceedings of JENC7, Budapest.

[Linn 88] J. Linn (1988). *Privacy enhancement for Internet electronic mail: Part I: Message encipherment and authentication procedures*, RFC1040.

[PGP 97] http://www.pgp.com

[Ragget 96] D. Ragget (1996). *HTML 3.2 Reference Specification*, 1996, http://www.w3.org/pub/WWW/TR/WD-html32

[Secret 94] A. Secret (1994). *Agora, for those who cannot be in the Arena*, 1994, http://www.w3.org/hypertext/WWW/MailRobot/send.html.

[Spencer 94] H. Spencer (1994). *News Article Format and Transmission*. RFC1036bis

[WebMail 95] http://www.best.com/~pierre/web-mail/

Author Information

Manfred Bogen has been active in the area of group communication, X.400 development, and X.400 standardization since 1983. In 1987 he became head of the VaS group being responsible for the provision of value-added services. He studied computer science at the University of Bonn and is co-author of two books about X.400 and distributed group communication. At present, he is the convenor of the TERENA working group on quality management for networking (WG-QMN) and a member of the TERENA Technical Committee and the Internet Society.

Guido Hansen has completed his vocational training as a mathematic-technical assistant (MTA) at GMD in 1995. During a practical work as part of his vocation he wrote the first version of W3Gate. He is now working in the VaS research group in the fields of W3Gate, WWW and Web-based projects with external co-operation partners from the media industry.

Michael Lenz received his master's degree in informatics (computer science) from the university of Bonn. He is working at GMD in the Department for Network Engineering since 1993. His major topics are value-added services, security and information systems. Since late 1993 he is involved with the establishment and maintenance of information services like WWW at GMD and within various external projects. He is especially engaged in W3Gate's security and QoS design.

The authors can be reached at GMD - German National Research Center for Information Technology, D-53754 Sankt Augustin.

Is the Web a Secure Environment for Electronic Commerce?

Manfred Bogen <Manfred.Bogen@gmd.de>
Michael Lenz <Michael.Lenz@gmd.de>
Andreas Reichpietsch <Andreas.Reichpietsch@gmd.de>
Peter Simons <Peter.Simons@gmd.de>

GMD - German National Research Center for Information Technology
D-53754 Sankt Augustin

Abstract: About 30 million people are using the Internet at present, and approximately one million new users log on each month. More and more business and electronic commerce is done through the Internet. What hinders the rest of the world to connect to the Internet and to use it commercially? Three reasons have to be mentioned: connectivity, costs, and last but not least anxieties.

This paper addresses the last issue and related remedies. It overviews conventional and extraordinary security threats and counter-measures especially related to the World Wide Web. It shows that security is relative and that the balance between convenience and protection is hard to find. In consequence, the Internet does not offer more security than normal life and one has to go a certain risk to benefit from its potential.

Keywords

World Wide Web (WWW), security, HTML, HTTP, cryptography, PGP, PEM, electronic commerce

Introduction

The Internet started in the academic community. It was used for scientific communication and cooperation, for the exchange of research results, and to address and discuss specific research subjects. Security was no issue at all. This changed with the Web and its applications. Electronic Commerce provides the capability to sell and buy products and information through the Internet with the major goal to attract new customers and to place a company in the new, challenging, worldwide market.

What do you need for business on the Web or electronic commerce? An attractive, up-to-date and state-of-the-art Web presence with feedback possibilities and interactive components is a base. A good ranking by the worldwide search engines and directories is crucial. Additionally, it makes sense to have advertisement for your Web presence in conventional media. You have to know whom your clients and your competitors are. In a next step you have to make yourself attractive for new advertisement by a clear marketing strategy. At last, you need an appropriate electronic payment method. Having implemented all this, you finally need regular customers who really trust in the electronic commerce idea, simply that it all works as well or even better as in normal life. This may be a problem mainly due to security considerations.

This paper addresses Web security. It is structured as follows: after the presentation of a [Conventional Set-up of a Web Service], [Additional Threats] describes what worst case can happen to a Web-based service. [Possible Technical Solutions & Legal, Social and End User Constraints] lists the things which one can do in principle. In its [Conclusion] the paper emphasizes again that Internet is as secure as normal life and that one has to go a certain risk to benefit from it.

Conventional Set-up of a Web Service

Four parties can be involved for electronic commerce: the content provider as source and owner of the data, the Web editor who produces HTML pages, the Web masters responsible for the installation and operation of the

Web server [Liu et al 94], and the public, the actual target group. All of them have access to the Web site with different rights.

[SECFAQ 97] identifies four overlapping types of risk in running a Web server:

- Private or confidential documents stored in the Web site's document tree falling into the hands of unauthorized individuals (unauthorized disclosure).
- Private or confidential information sent by the remote user to the server being intercepted.
- Information about the Web server's host machine leaking through, giving outsiders access to data that can potentially allow them to break the host.
- Bugs that allow outsiders to execute commands on the server's host machine, allowing them to modify and/or damage the system. This includes "denial of service" attacks, in which the attackers pummel the machine with so many requests that it is rendered effectively useless.

To cope with the existing rather conventional threats a combination of traditional host and network security techniques has to be applied [Fig. 1]. The Web serve, accessible from the Internet and the Intranet through a conventional TCP/IP-based network, is placed on a secure server net together with other services like email, database or news. It is protected by a firewall.

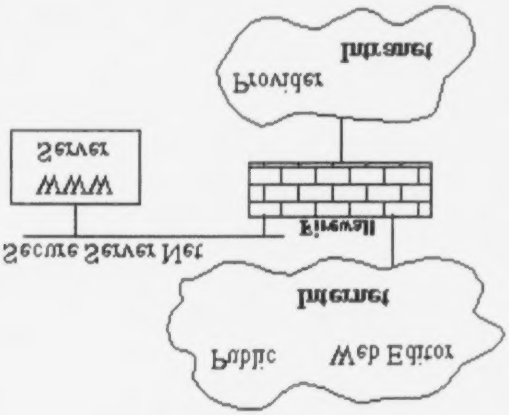


Figure 1: Conventional Web Set-up

To restrict the public read access to Web documents, Web servers themselves additionally offer two different methods. The first one is based on the Internet address of the requester; the second one is based on normal user name and password access authentication for individuals or even groups. The access control and user authentication can relate to the server in total or to respective directories.

Using secure logins [SSH 97] prevents passwords from being transferred as clear text over the network. Password cracking programs as crack [CRACK 97] can be used to detect weak passwords. Local access restrictions (like UNIX users' and group rights) can be used to ensure that only authorized staff, e.g. the Web editors, can modify the Web documents. With tools as tripwire [TRIPWIRE 96] any illegal modification can be detected. As a complement, virtual private network (VPN) techniques, offered by modern firewalls, can be used to protect unintentional modifications of the Web pages by the Web masters.

On a regular basis Computer Emergency Response Teams [CERT 97] all over the world announce known vulnerabilities in available software and give advice. With respect to the Web, only servers without known security bugs should be used and they should run with minimal privileges. Also all CGI scripts [CGI 95] should be scanned very carefully for potential security holes as they are called by the Web server and hence inherit its access rights.

Additional Threats

All security efforts described above concentrate on defending the Web site itself. There is no defense against more sophisticated attacks, though, which can be performed by any individual that has authorized or unauthorized access to central routers, a cache, or a Web server. All attacks described below can be classified as "man-in-the-middle" attacks [Fig. 2] and are very dangerous, because they are fairly easy to perform, but very hard to detect.

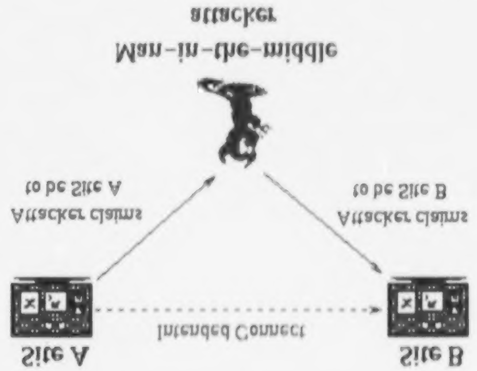


Figure 2: Man-in-the-middle Attack

Redirection of IP Packets

IP routers can redirect IP packets to an address different from the original destination. Most HTTP demons can act as a virtual host. So, it is possible for an Internet Service Provider (ISP), to redirect all packets originally addressed to 'www.netscape.com port 80' to an entirely different machine with a WWW server pretending to be 'www.netscape.com'. Any user accessing the address 'www.netscape.com' via HTTP through the modified router will now be talking to the fake server. Each incoming CONNECT package is answered by the fake server. Calls to CGI programs are forwarded to the original site, and the result is then sent to the site which originated the CONNECT. The fake server, though, may modify the results before sending them back.

There are many advantages for an attacker in this set-up. He can gather user passwords, distribute software with backdoors, or simply put material on the fake server that damages the credibility of the vendor, whose site is "attacked". There is virtually no countermeasure against this kind of attack. Even worse, the original server does not even notice the attacks. Users, who have their packets redirected to a different machine, have almost no means of detecting it, and even if the attack is detected, it is very difficult to find the machine that does it.

Web Spoofing

While surfing through the Web, people visit numerous servers all over the world. The Web documents contain references to related information on other servers. By clicking on such a hyper link, the browser will contact that server directly or indirectly (cache server) and request the document. In a Web spoofing attack [Felten et al 96], the attacker sets up a Web site containing links to numerous other servers. He tries to attract users to his pages and to follow the links that all lead to a special script on the attacking site. It fetches the respective document from the original site and returns it to the requesting site after all links are manipulated in a way that they also call this little script ('masquerade').

Some additional provisions are necessary to complete the illusion. Normally, the URL of the current document is shown within the browser as well as the URL of the corresponding document while the mouse cursor is over a link. By adding some JavaScript to the manipulated documents this can be achieved very easily. For the users nothing has changed. At a first glance they receive the requested documents. Unfortunately, all network traffic goes through the attacking site; e.g. all transactions can be scanned and potentially changed.

This attack however has its weaknesses. The victims may leave the faked Web server by selecting a bookmark or a hot list entry or by entering an URL within the browser's open link menu. In some respect this attack is easier to detect than a router manipulation: The browser's 'view document info' menu may reveal the real, i.e. rewritten URL of the current document. Also viewing the document source shows the manipulated links. Unfortunately the HTML source of a document is not easy to read especially for a novice user.

On the other hand this attack might be more dangerous to some extent. Anybody can undertake it and after visiting the attacker's site all subsequently visited documents and Web sites are under attack and might be manipulated. To cover the tracks, this attack will in general be combined with breaking into someone else's Web site.

WWW Cache --- Do Not Trust (your) Cache Server

WWW caches [Cormack 96] are used to optimize Web access and to minimize transfer costs. There are local caches on personal computers and central cache servers on entire networks or with ISPs. Enabling cache servers in the browser indicates: there is no guarantee that the page comes from the indicated server. The person who is controlling a cache server is able to change the cached pages. Using cascaded or cooperating Web caches makes this danger even bigger. Internet Service Providers may redirect all Web requests (i.e. requests on port 80) to their cache server, even if the user disabled cache usage.

Direct attacks against a cache server will effect much more users than attacks against single computers in the network. One could resolve the requests of a cache server with faked information. A potential attack on a cache server is to change the time stamps of the cached documents or to change the system time on the machine. Updates remain undone, even if the original document is changed and the requester gets the old (wrong) version of the document. Man-in-the-middle attacks are of special importance, because a cache server always is a 'man in the middle'. Standard tests for such attacks on both ends do not work and are often disabled.

Using a cache server to restrict public access to company-internal Web pages may cause problems. For an internal Web server document requests now come from the cache server and not from the outside, so it answers them. The same problem occurs if distributed caches are used and intra-domain cache servers cooperate with cache servers outside. Cooperation between databases and cache servers will cause problems if restricted pages tunnel through the database or the database is insecure or misconfigured.

Attacking Cryptographic Site Authentication

Many sites try to address the authenticity problem by using cryptography. The most popular method today is the Secure Socket Layer [SSL 96], which protects the server and the client in two ways: the peers of a CONNECT are authenticated via digital signatures, and the data transfer itself is encrypted. In theory, this would make any of the attacks mentioned above infeasible, but in practice, there are still a number of open doors in this set-up.

For one, the cryptographic keys are not strong enough. Export versions of the Netscape browser, for example, are restricted to 40-bit cryptography due to the American ITAR regulations. Cracking a 40-bit with a simple brute force attack is far from being impossible, though. Just recently a group of students broke a 40-bit RC5-key within 4 hours [RSA 97]. A large company or organization with enough computing power is able to achieve the same result even faster.

Breaking the encryption is not even necessary to attack a server, though, because everybody is able to fake the required keys. A server acting as "man in the middle“ could simply answer with its own set of keys. The chance that anybody would notice it are very small, because there is no reliable way to check the authenticity of the keys that are used to guarantee the authenticity of the server.

Last, but not least, the fake server could simply drop the whole SSL business completely. SSL protected data transfers are very uncommon in the Web, even today. If you do not know that the original server supports SSL, you will not miss it while talking to the fake server anyway.

Possible Technical Solutions & Legal, Social and End User Constraints

The only defense against a "man in the middle" attack is to verify that you are connected to the machine you expected to answer your CONNECT. This can be achieved by using digital signatures. The digital signature technology is known for more than two decades now and it has proven to be reliable. Furthermore, encryption of transferred data must become the standard --- not the exception.

Technical Constraints

The digital signature that certificates the authenticity of the server can easily be faked, as mentioned above. As a result, the keys that issued the signature have to be certificated themselves, too. How do you verify the keys that certificated the keys? A reliable key exchange through the Internet is not possible, because the key exchange is subject to attacks, too.

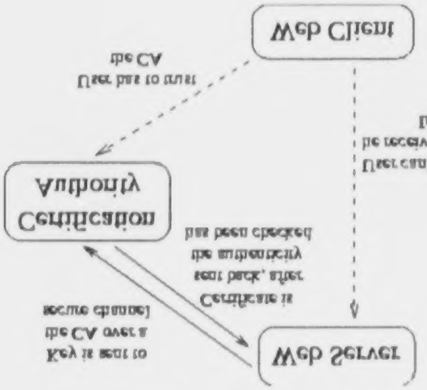

Web Client
User has to trust the CA
User can't trust the key he receives through the Internet
Certification Authority
Certificate is sent back, after the authenticity has been checked
Key is sent to the CA over a secure channel
Web Server

Figure 3: Web of Trust

The only way keys can be exchanged securely is over a secure channel. Establishing a secure channel between two parties is very difficult though; it may even be impossible if the peers live in two different parts of the world. Various concepts have been introduced to solve this problem: "The Web of trust" [PGP 97, Fig. 3] or central certification authorities as used by PEM [RFC1040 88]. Neither approach has solved the problem of authenticity in the Internet satisfactorily on a global scale, though.

Social Constraints

Many countries have laws that limit the usage of all kind of encryption. The United States do not allow strong cryptography to be exported. Thus, all software exported from the US is only capable of handling encryption keys up to 40-bit. This is too weak for today's computing power. Other countries like France have banned encryption completely because they are afraid of criminals or terrorists abusing this technology. Some more countries are currently evaluating whether encryption should be outlawed.

End User Constraints

The Internet exists for more than twenty years now, but it did not become that popular until an easy to use graphical user interface was available: the WWW. If electronic commerce wants to be successful, it has to attract not only the computer literate population, but also people who do not know very much about computers, networking, and cryptography.

This is a serious limitation in terms of what can be done, because security and comfortable usage usually contradict each other. The end user does not want to be bothered with complicated authentication protocols. He does not want to memorize a dozen unique passwords or PINs.

Conclusion

Cryptographic authentication of WWW sites makes an attack much more difficult, but not impossible. Until cryptographic authentication is deployed by a much larger number of servers than today, it is practically ineffective. The very nature of a cooperating open network makes it impossible to guarantee complete security and privacy. Every system in use can be abused. Though being technically possible to ensure authenticity of connecting peers and security of transferred data, the required effort would by far outweigh the benefits.

It is a common misunderstanding that the Internet is something revolutionary new. The Internet is a new technology to do things in a much more efficient way than in the past. So the same security standards should be applied, as in the rest of the business world. A mail order vendor, for example, has no means of verifying every phone call and every facsimile he receives. Nonetheless he accepts the risk of received faked orders, because the profit he makes by real orders warrants the costs abuse causes.

When using the Internet for electronic commerce, one has to take risks too. However, the advantages of the Web for electronic commerce remain: economic, worldwide and around the clock. The drawbacks can not be remedied completely. Instead of worrying about obscure attacks, possible abuse, and all kind of risks, we recommend to inform about all aspects of electronic commerce in the Web in time and honestly. This was the purpose of our paper. Finally, we can answer our main question whether the Web is a secure environment for electronic commerce like Radio Erivan: yes, in principle.

References

[CERT 97] http://www.first.org/

[CGI 95] http://hoohoo.ncsa.uiuc.edu/cgi/security.html

[Cormack 96] Andrew Cormack, Web Caching, A Report, University of Wales, Cardiff, September 1996

[CRACK 97] ftp://coast.cs.purdue.edu/pub/tools/unix/crack/

[Felten et al 96] Edward W. Felten, Dirk Balfanz, Drew Dean, and Dan S. Wallach, 'Web Spoofing: An Internet Con Game', Technical Report 540-96, Department of Computer Science, Princeton University

[GS 97] Simson Garfinkel and Gene Spafford, 'Web Security & Commerce', Sebastopol, CA: O,Reilly & Associates Inc.

[Liu et al 94] Liu, C., Peek, J., Jones, R., Buus B. and Nye, A. (1994). Managing Internet Information Services. Sebastopol, CA: O,Reilly & Associates Inc.

[PGP 97] http://www.pgp.com/

[RFC1040 88] J. Linn, "Privacy enhancement for Internet electronic mail: Part I: Message encipherment and authentication procedures"-, 01/01/1988

[RSA 97] http://www.rsa.com/rsalabs/97challenge/status.htm

[SECFAQ 97] http://www-genome.wi.mit.edu/WWW/faqs/www-security-faq.html

[SSH 97] http://www.cs.hut.fi/ssh/

[SSL 96] http://home.netscape.com/assist/security/ssl/protocol.html/

[TRIPWIRE 96] ftp://coast.cs.purdue.edu/pub/COAST/Tripwire

Author Information

Manfred Bogen has been active in the area of group communication, X.400 development, and X.400 standardization since 1983. In 1987 he became head of the VaS group being responsible for the provision of value-added services. He studied computer science at the University of Bonn and among other scientific publications he is the co-author of two books about X.400 and distributed group communication. At present, he is the convenor of the TERENA working group on quality management for networking (WG-QMN) and a member of the TERENA Technical Committee.

Michael Lenz received his master's degree in informatics (computer science) from the university of Bonn. He is working at GMD since 1993. His major research topics are value-added services, security and information systems. Since late 1993 he is involved with the establishment and maintenance of information services like WWW at GMD and within various external projects.

Andreas Reichpietsch joined GMD in 1985. He was working in operating system research. Since 1992 he works as system and network administrator in the Department for Network Engineering. Special interests are computer- and network-security and applied cryptography.

Peter Simons is a student of computer science at the University of Bonn who works part time for the GMD's Media Communication Institute. His special interests since 1989 are computer security and cryptography in networking environments. He is author of several publications and a book about the information highway and related problems.

Information Brokering on the World Wide Web

Stéphane Bressan and Thomas Lee
MIT Sloan School of Management
E53-320, 50 Memorial Drive
Cambridge, MA 02139, USA
{sbressan,tlee}@mit.edu

Abstract: Information Brokering is the process of collecting and re-distributing information. As the rich but unstructured sea of data that comprises the World Wide Web continues to grow, so too will the demand for information brokering services. This paper outlines a methodology and an architecture to support information brokering on the Web. Specifically, the two innovations which are described facilitate the construction and distribution of customized views that integrate data from a number of Web sources. An original technique is presented for the extraction of data from the Web and a solution based on JavaScript and HTML is presented to support the embedding of SQL queries within Web documents.

1 Introduction

In [Levine 95], the author defines Information Brokering as the *"business of buying and selling information as a commodity"*. Levine tracks the modern origin of information brokering to the French in 1935. The French SVP was an organization supplying information on demand over the telephone. This paper refines the definition of information brokering to mean the process of collecting and re-distributing information where Information Brokers are organizations which supply brokering services [Fig. 1].

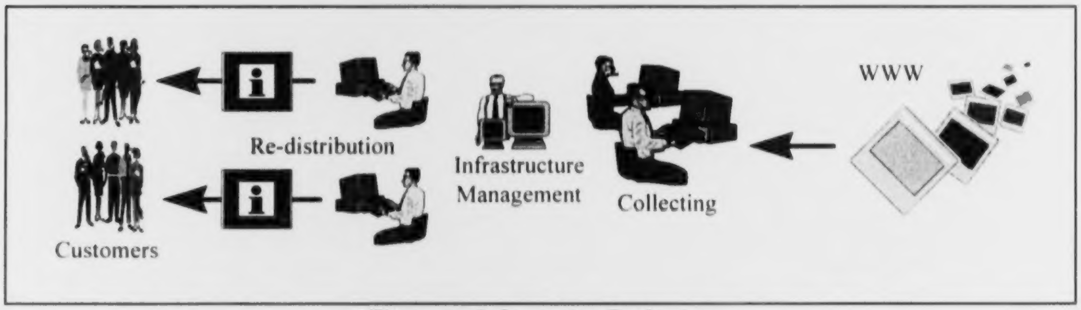


Figure 1: Information Brokering

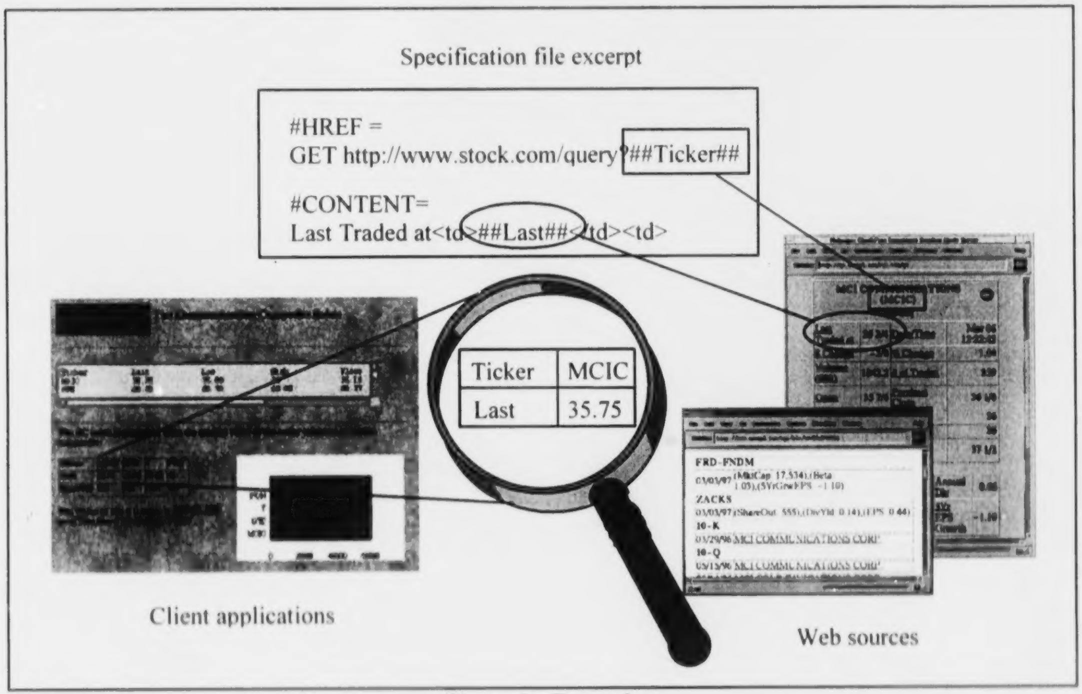
Many projects [Garcia-Molina 95, Levy et al. 95, Fikes 96, Martin 96], including the COntext INterchange (COIN) project [Bressan et al. 97a, Bressan et al. 97b] from which this work stems, as well as research programs like the American DARPA I3, or the European Esprit and Telematics programs, focus on the general issue of information integration. In particular, the above referenced projects and programs leverage the Mediation reference architecture presented in [Wiederhold 92]. Although COIN addresses the general issue of semantic integration of heterogeneous information systems, this paper focuses on the specific issue of the collection and re-distribution of information on the World Wide Web in Internet based public or corporate information infrastructures.

Consumers today often have specific information needs, which are satisfied through the aggregation and analysis of individual data sets. While the World Wide Web offers a tremendously rich source of *data*, it fails to satisfy a user's *information needs* in at least two ways. First, information providers are constrained in their ability to flexibly present and represent data to end-users. Users may be interested in graphical rather than numeric representations or aggregations rather than raw values. Providers, however, face a trade-off between flexibility and security. Existing tools sacrifice expressiveness in exchange for guarantees about more limited behavior. Second, exacerbating the problem of representation is the challenge posed by extracting information from heterogeneous sources. Because information accessible via the Web ranges from free text to well-structured tables, users lack a uniform means for isolating and retrieving information of interest from distinct Web pages. Semi-structured sources, such as HTML tables, mimic relational structures without similar guarantees of relational behavior.

Consider, for example, the case of two financial analysts who are assessing the current state of the United States residential, long distance telephony industry. One analyst might like to graphically view the performance of each player in the residential, long-distance telephony market with respect to a "Communications industry index" calculated as an aggregation of share price and shares outstanding for all market actors. A second analyst might be interested only in viewing raw figures such as closing stock price. [Fig. 2].

As illustrated in the right-hand side of [Fig. 2], however, the data may only be available over the Web in a disaggregated form. Furthermore, the data may be indexed by company name rather than by industry sector, and may be distributed over multiple Web pages. Finally, for a specific page, users have no mechanism for identifying which values are of interest.

Information brokering is the process of identifying a user's *information needs*, *collecting* the requisite *data* from disparate sources, and then *redistributing* the formatted and processed data to the consumer. News agencies, Chambers of Commerce, and financial analysts are all examples of institutions who engage in some form of information brokering. In the example above, the analysts might perform the brokering task themselves or they might rely upon a third-party service (or internal department) to collect and compile the requisite data.



```
#HREF =
GET http://www.stock.com/query?##Ticker##

#CONTENT=
Last Traded at<td>##Last##</td><td>
```

Ticker	MCIC
Last	35.75

Figure 2: Example

This paper presents an architecture for Web-based information brokering. The Web brokering task is separated into three functions: collecting, redistribution, and infrastructure management. The relational data model [Ullman 88] is interposed as an abstraction between the different tasks; therefore, queries are posed in a uniform manner and results may be reduced to a standard format. In [Section 2], the architecture is developed by expanding upon each of the three brokering sub-tasks. [Section 3] discusses the motivation for this paper's approach and comments on related work and future directions. The paper concludes by speculating on the role of information brokering within the more general context of heterogeneous information integration.

BEST COPY AVAILABLE

2 Broker architecture

As illustrated in [Fig. 3], an architecture to support information brokering may be constructed around the subdivision of the brokering process into collecting and re-distribution. Whether within a corporate intranet or over the open Internet, Web brokering introduces a third component, infrastructure management, that supports the collecting and redistribution sub-processes.

2.1 Collecting

Web wrapping is the process of collecting or extracting data from web documents and of structuring the data into a relational form. A wrapper is a software component that serves as a gateway between the client applications and the World Wide Web. It exports relational views of some selected information on the pages, accepts SQL queries against this schema, and extracts, formats and returns the data in a relational table.

For the purposes of this text, a document is the value returned by a Hyper Text Transfer Protocol (HTTP) request. A document is retrieved by selecting a method (usually POST or GET) and a Uniform Resource Locator (URL). A URL (e.g. http://www.stock.com/query?MCIC) specifies a protocol (http), a server (www.stock.com), the path on the server (/query), and, optionally, a parameter string (?MCIC). Documents corresponding to a given method and URL may be dynamically generated or may vary over time as contents are updated.

In general, the automatic extraction of data from a document is difficult. The document may not have any identifiable structure. We are here considering categories of documents which contain some observable structure such as, for instance, Hyper Text Markup Language (HTML) tables or lists that we expect to remain though some of the content varies. Data can be extracted from such documents, if the structure is known in advance, using pattern descriptions and pattern matching. Although we are currently working on techniques combining parsing of the HTML structure and regular expression pattern matching, we will only present in this paper the technique we have already implemented based on the sole pattern matching of regular expressions.

In the example of [Fig. 2], today's lowest price of the MCIC security is in a table cell immediately after a cell containing the string "Day Low". The regular expression pattern (in the Perl syntax) "Day Low.*</td><td>(.*)</td>" matches the sub-string "Day Low</td><td>35 5/8</td>" in the document source. It binds a variable (corresponding to the sub expression in parenthesis) with the value "35 5/8". In a similar way we can match other data from the document such as the last price of the security, the highest price during the day, or the price at the previous close. A single regular expression can bind more than one variable and we can define more than one regular expression for a given document.

It is interesting to use such a description of the content of a document if we expect the document content to vary inside the limits defined by the identified structure. Such a situation is common on the Web today. However documents vary in two dimensions: over time as their content is updated, but also as the URL varies. The latter case typically corresponds to documents automatically generated by programs called via the Common Gateway Interface (cgi) for which the object body of the URL, the parameters, can change. Here, also we use the same pattern matching technique to characterize the variations in the URL. In our example, the Ticker (the identifier of the company in the stock exchange listing: MCIC, T, FON, GT) is part of the URL.

We call a page the set of documents defined by a URL pattern. A page specification contains a list of regular expressions defining the data elements in the documents. The reader notices that a regular expression can find several alternative bindings on the same document.

We aim to collect the data from Web documents into a relational format. To each variable in the regular expressions, we associate an attribute of a relation. We rewrite the regular expressions using the attribute names (e.g. "Day Low.*</td><td>##Low##</td>" or "http://www.stock.com/query?ticker=##Ticker##", where ##attribute## identifies the attribute in the expression).

The relation is defined by the process of collecting data from the documents corresponding to the page specification. Alternative bindings on a document are collected into tables. The results for each individual documents are collected in the union of all the tables.

In our example we have defined a view with the schema (Ticker, Last, Low, High, Close). One tuple in that view is (MCIC, 35 3/4, 35 5/8, 36, 36 1/8). The view contains all the corresponding tuples for each company ticker.

We observed that the data one is interested in is often spread over multiple documents corresponding to more than one page (i.e. with various set of patterns and URL patterns). In our example, the number of outstanding shares for a given ticker can be obtained from a different document. A relation corresponds to a set of page definitions for the set of documents that one needs to access to collect all of the data. The relation is defined as the natural join of the views for each page (i.e. the join over identically named attributes). The number of outstanding shares and the other data collected on the first page of our example are joined on the ticker's value. If we call r1 and r2 the views corresponding to the two pages, the relation (call it s) is defined as a view:

```
DEFINE s AS SELECT r1.Ticker, r1.Last, r1.Low, r1.High, r1.Close, r2.Share
FROM r1, r2 WHERE r1.Ticker = r2.Ticker;
```

In other words, relations are defined as views under the Universal Relation concept [Ullman 88]. *Attributes of the same Name are the same*. We have chosen a *window function* based on the systematic natural join, which has a clear semantic but may lead to inefficient evaluations. Alternative window functions can be preferred.

In many practical cases, in order to keep the relationship among the different pages, we may need to use ancillary attributes (codes, html file names, values of menus in forms, etc) which are not relevant from the application point of view. For this reason we associate which each relation definition, an export schema which corresponds to the attributes visible from the application. The definition of each relation: its name, attributes, export schema, and set of page definitions (URL pattern and regular expressions) are grouped into a unit we call the specification file.

Finally, the different specification files defining a set of relations are the parameters of the wrapper program. Given a query on the exported schemas and the specifications, the wrapper generates, optimizes, and evaluates a query execution plan. It combines the result of the query in the form of a table accompanied with additional information such as the name of the attributes, the number of answers, and administrative data of potential use for the application (such as time stamps).

2.2 Redistribution

Redistribution involves posing queries, integrating the data retrieved from one or more Web wrappers, and formatting the data to meet a client application's requirements. Redistribution may also require additional data processing. In the earlier example, the financial analyst calculating the "Communications industry index," stock values from all residential, long-distance market actors are aggregated. The solution introduced here leverages "server side includes" (SSI) (It can also be implemented using CGI provided that care is taken to avoid the parsing of the entire HTML document) and JavaScript to submit queries and to process the results.

Queries are embedded within Web documents and submitted via SSI. A single command both defines the query and declares a handle on the result. For the telecommunications industry analysis referenced above, the query might appear as (in SSI syntax):

```
<!-- #exec cmd="wrapper query=Select Ticker, Share, Last From s Where Ticker in (T,MCIC,FON,GTE);
handle=tel_index"-->
```

When a client application requests the document containing a query, the Web server invokes the query and returns to the client a document where each command line is replaced by the query result.

Rather than returning query results as HTML formatted text which is directly substituted into a Web document, the SSI introduced in this paper returns a JavaScript program. The JavaScript program defines a JavaScript object which is referenced by the handle in the SSI declaration. The result object contains a query result table, attribute names, data types, and administrative information provided by the wrapper such as time-stamps. Values are accessed and formatted from within the source HTML document by calling primitive or advanced JavaScript functions to the JavaScript result object. Basic functions are provided as direct methods of the result object, and advanced functions could be defined by Web page designers or loaded via SSI from libraries.

As illustrated in the left-hand side of [Fig. 2], the combination of SSI and JavaScript demonstrated in this paper offers tremendous flexibility with respect to both data presentation and data re-use. Values may be aggregated or displayed in raw form. Data may be formatted in tables or graphs. The Communications Composite Index in [Fig. 2] is generated by the following HTML-embedded JavaScript program:

```
<SCRIPT>
index=0; for (i=1;i<size(tel_index);i++) {index=tel_index[i][2]*tel_index[i][3];}
document.writeln(index);
</SCRIPT>
```

Data from a single result object may be reused in multiple portions of a document without re-submitting the same query or a sub-query (SSI). For example, the bar chart on the left-hand side of [Fig. 2] was generated by passing the result-object to a public domain Java applet. In summary, we propose a protocol based on SSI and JavaScript, which offers maximum flexibility in presentation and ex-post data manipulation (particularly when extensions of HTML such as Dynamic HTML are used) while minimizing the number of external calls and leaving much of the formatting to the client application (JavaScript).

2.3 Infrastructure Management

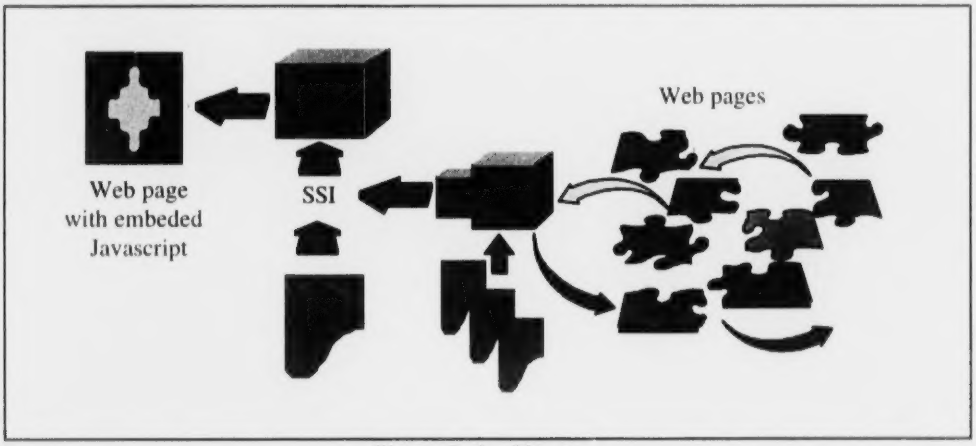


Figure 3: Architecture

[Fig. 3] summarizes the architecture we propose. The structure of the information located on various disparate Web pages is described in the specification files. The wrappers receiving a query extract, combine, and as required by the query and the relational schema. Queries are inserted as SSI-commands in the HTML source. The commands are processed the wrappers on request of the Web server. The result of the SSI-processing is a JavaScript object available on the document served. The data are displayed in different formats and combinations by means of JavaScript functions embedded in the HTML pages and processed on the client machine. Different designers can freely combine the data available in appropriate formats on their Web pages or prepare reusable components using Dynamic HTML layers.

3 Discussion and conclusion

3.1 Related Work

Wrapping

There currently exist a number of broader information integration efforts that employ some form of non-intrusive wrapper for exporting a uniform query interface to and extracting data from disparate sources. Wrapper technologies differ with respect to the query language interface, the kinds of sources supported, and the functionality delegated to wrappers. TSIMMIS [Garcia-Molina 95] wrappers process Mediator Specification Language queries on OEM objects. TSIMMIS wrappers, which map sources into OEM objects and create optimized query execution plans, have been demonstrated for the LORE database and network information services like finger and whois. SIMS leverages a LOOM knowledge representation to wrap Oracle databases and LOOM knowledge bases by mapping data sources to subclasses of a LOOM ontology. Though queries are posed as LOOM statements, the wrapper itself mainly relies upon other components of the SIMS system for query processing.

BEST COPY AVAILABLE

Redistribution

Redistribution is a particularly challenging problem for information brokers because it is difficult to determine a priori the dimensions of a consumer's demands. Even within a single institution, as in the case of the two telecommunications financial analysts above or across a corporate, departments and divisions may present different aggregation and formatting requirements. Conventional solutions rely either upon CGI scripts or SSI. CGI solutions are overloaded so that scripts not only request and retrieve data but also format the resulting data within a Web document. Such an architecture is a drawback when the same data is re-used in more than one application. The proliferation of scripts complicates quality control and compounds the information broker's security vulnerabilities due to multiple, repeated gateways between the outside world and the broker's servers. SSI solutions also typically overload a single operation with query request, retrieval, and format. SSI calls typically return HTML to facilitate direct substitution into the encompassing HTML. Moreover, as alluded to above, SSIs can introduce querying inefficiencies. Rather than submitting a single query to an external source and then caching the results (JavaScript object or CGI variables), multiple SSI invocations are required. Even re-using the same data requires a new call. Finally, SSIs alone support only limited formatting capabilities. Standard extensions call for passing formatting parameters in the SSI ultimately reducing the call to a general scripting language.

3.2 Limitations and Future Work

As detailed, the current information broker is limited in three dimensions. First, the absence of a means to express knowledge about the page's content and the lack of such information limits the opportunities for optimizing the process of data collection by selecting the most appropriate documents. Second, although the combination of SSI and JavaScript affords great flexibility in data formatting and presentation, client applications are ultimately limited to rendering the JavaScripts and HTML returned to them by information brokers. A logical extension is to therefore enable clients to dynamically reformat values. A related limitation and attendant extension is the ability for end users to parameterize queries. In the current architecture, because queries are invoked from the information broker's Web server through an SSI, queries are articulated in advance and maintained as static pages by the broker. The third dimension of future work calls for enabling client applications to dynamically structure and submit queries through the information broker directly to wrappers, using a gateway (a Java applet) monitored from the client and collaborating with client-side JavaScripts. The described architecture has been implemented and is currently deployed within the Context Interchange system.

4. References

[Bressan et al. 97a] Bressan,S., Fynn, K., Goh, C., Madnick, S., Pena, T., & Siegel, M. (1997). *Overview of a Prolog Implementation of the Context Interchange Mediator.* Proc. of the Intl. Conf. on Practical Applications of Prolog.

[Bressan et al. 97b] Bressan, S. Fynn, K., Goh, C., Madnick, S., Pena, T., & Siegel, M. (1997). The Context Interchange Mediator Prototype. Proc of SIGMOD97.

[Fikes 96] Fikes, R. (1996). *Network based Information Brokering.* http://www--ksl.stanford.edu/kst/info-broker.html.

[Garcia-Molina 95] Garcia-Molina, H. (1995). *The TSIMMIS Approach to Mediation: Data Models and Languages.* Proc. of the Conf. on Next Generation Information Technologies and Systems.

[Levine 95] Levine, M. (1995). *A Brief History of Information Brokering.* American Society for Information System Bulletin,. February, 1995.

[Levy et al. 95] Levy, A., Srivastava, D., & Kirk, T. (1995). *Data Model and Query Evaluation in Global Information Systems.* J. of Intelligent Information Systems.

[Martin 96] Martin, D. (1996). *The Information Broker Project.* http://www.ai.sri.com/~martin/broker/techrep/memo.ps.gz

[Schum 96] Schum, A. (1996). Open Database Connectivity of the Context Interchange System. MIT Master thesis. Dpt. of Elect. Eng.

[Tomasic et al. 95] Tomasic, A., Rashid, L., & Valduriez, P. (1995). *Scaling Heterogeneous databases and the Design of DISCO.* Proc. of the Intl. Conf. on Distributed Computing Systems.

[Ullman 88] Ullman, J. (1988). *Principles of Database and Knowledge-base Systems, Volume 1* . Computer Science Press, Rockville, MD.

[Wiedehold 92] Wiederhold, G. (1992). *Mediation in the Architecture of Future Information Systems.* Computer, 23(3).

Teams, Tasks, and Notices:
Managing Collaboration via the World Wide Web

Charles L. Brooks
Frederick J. Hirsch
W. Scott Meeks

The Open Group Research Institute, Cambridge, MA 02142 USA
{c.brooks,f.hirsch,s.meeks}@opengroup.org

Abstract: The World Wide Web is advancing from a "read-only" medium to one supporting local publishing, distributed authoring, and soon, distributed collaboration. In this paper, we present an experimental system we have developed that incorporates novel ideas of notification, visualization, and extended Web services to support the creation, management, and effective use of collaborative workspaces in the World Wide Web.

1. Introduction

As originally conceived by Tim Berners-Lee in the early 1990s, the World Wide Web supported the creation and publication of content as well as the reading of that content. This support has not been generally available because most HTTP server vendors do not directly implement the HTTP PUT method [Apache 97], but during the past year several commercial vendors have offered Web products that permit the remote authoring of content, such as Microsoft's FrontPage editor [Frontpage 97] and Netscape's Composer HTML editor [Composer 97]. In addition, systems like BSCW [Bentley 97] have publicized the notion of *workspaces*, where remote collaborators can publish, review, and critique the work of other members of their project.

Collaborative workspaces are especially effective in an Intranet environment, where Internet technology is used in the context of a single organization, such as a particular company or university. Restricting this environment to more well-defined organizations allows individual users to be known, policies to be set, and consistent administration of the Web to be performed. A shared workspace in an Intranet environment thus can take advantage of these restrictions to provide a more effective environment for information sharing.

One important class of business objects is project documents representing either individual project deliverables or sets of tasks. These documents may be registered into the shared workspace at any time and assigned to categories to make retrieval easier. Each document has an owner, but may be passed from individual to individual for modification. Documents have states, such as created, revised and sealed (no changes permitted). Documents also have subscription lists of users associated with them, permitting notifications to be sent to interested parties when document state changes. Subscribing parties may be either individuals or applications, and these parties may take various actions upon receiving these notifications.

It is this notion of asynchronous notification that ultimately supports the ability to perform cooperative work within this workspace. As an example, consider a typical review and authorization process. A special reviewers list is generated for each document, which itself is associated with a “document creation” task. When this task is completed, the draft document is marked as submitted, and each reviewer is notified. Reviewers indicate approval by generating an approval notice. Once all reviewers have approved the document, the task is now marked as completed and the document may be sealed (indicating no further changes), an action which generates another round of notifications.

2. Scenario

The environment described above is suitable for many different team organizations, including both business and education. A simple scenario (drawn from the education domain) both illuminates and motivates our work. This scenario follows three students as they work to complete their term project in a class on mobile distributed computing.

Alice, Bob, and Carol are members of a class on mobile distributed computing using Java. The class has a Web site which has documents and student information managed by an extended Web service called the Mediator [Mediator 96]. This Web site contains readings for the class, assignments, and information about individual students and project groups; it is also where students deposit their homework. The Mediator provides access control such that class readings and assignments can be read by all the students but only modified by the instructor, and homework can only be read by an individual student and the instructor. Whenever a document is added or modified within the document space controlled by the Mediator, a notification is sent to anyone who has subscribed to these Mediator notifications. The student's HistoryGraph applications (see [Desktop Applications]) can be configured to receive these notifications and display them by adding or modifying the node in the display corresponding to the document.

The instructor assigns the final project for the class by creating a document describing the project that involves forming a team of 3 to 4 students to design and implement a Java applet and associated classes to provide a "guided tour" of the Web. Alice, Bob, and Carol have already been studying together for the class and decide to form a team for the final project. They use the Mediator to register their new team: this sends out a notification that informs the instructor that they will be working together, and she approves their decision by generating an approval notice for their team document. First the team has to design the architecture of their applet. They record the results of their research and design as a tree of URLs, and they save this tree of URLs via the Mediator so that they can refer to it during the rest of the design and implementation process as well as include it in their final report.

They then begin the implementation phase. The code they write goes into documents that they save via the Mediator, sending out notifications so that each one can monitor progress of the others. By using the notification recording facility, they can even coordinate their work when not logged in. So, for example, when Alice goes away for the weekend, she invokes the "Record Notices" function before she leaves. Carol comes in over the weekend and completes an important user interface class that Alice needs. When Alice begins working on the project again on Monday, she invokes the "Retrieve Notices" function, retrieving any stored notifications including the notification about the new UI component, which she sees that she can now begin using.

Finally, they are done implementing and testing and have to deliver the project to the instructor. They create a tree of URLs providing a "guided tour" of the project which they save via the Mediator, and then mark all their pieces as done. The Mediator automatically sends out a notification which lets the instructor know when everything is complete. Grading is straightforward. The instructor follows the tour and verifies that the applet meets all of the requirements. She adds some annotations noting some elegant features in the design, and records their final grade as an annotation to the team's task document (readable and writable only by the team and their instructor).

3. Architecture, Design and Implementation

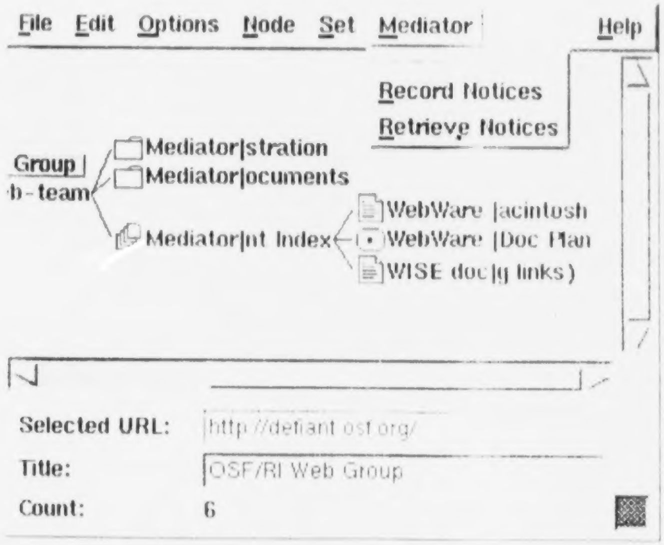
The previous scenario described how a group of individuals could function effectively as a team in an Intranet Web environment. This environment implements a system architecture consisting of desktop applications, group services and a notification framework.

3.1. Desktop Applications

Our desktop applications consist primarily of browsing associates: small, simple applications (compared to Web browsers and servers) which are not coupled to particular HTTP streams and can independently and asynchronously access the Web on the user's behalf. Our associates often take advantage of browser interfaces to observe user browsing actions, but this is not a requirement.

Each team member generally works with a number of Intranet Web pages, and, in addition to viewing individual pages, will want to visualize the group of pages as a whole in a manner which makes sense to them. The Web Activity Visualization System allows a user to see a graphical tree representation of the portions of Web sites she has visited and which is of interest. The user may manipulate this tree, work with the pages shown in the tree, and receive notifications in this visualization. We have implemented a prototype of this system called HistoryGraph [Hirsch 97]. HistoryGraph may be used in conjunction with our WhatsNew browsing associate [Brooks 95] for monitoring changes in the Web.

The main HistoryGraph display consists of nodes representing visited URLs and links representing the order in which URLs were visited. Only the first visit to a page is reflected within the tree. The nodes consist of small icons followed by an elided title or URL. The standard icon is a simple file folder icon and indicates no additional information about the page. Additional icons are used for nodes with additional information, such as the document icon (🗎) which shows a document which is stored in our Mediated access service, and the stack of documents icon (🗐) for an index of Mediated documents. The "sealed" icon (⊙) indicates that the document has been sealed and can no longer be modified (the icon is meant to represent an old-fashioned sealing wax seal).



Selected URL: http://defiant.osf.org/

Title: OSF/RI Web Group

Count: 6

Figure 1: Sample HistoryGraph Screen Showing Mediated Documents and the Mediator Menu

Along with a history mechanism, HistoryGraph provides a means for using and manipulating the visualization. Nodes may be rearranged by dragging with the mouse, or nodes may be removed so that the resulting representation is more meaningful to the user. Each node represents a Web resource, and has a URL and title associated with it as well as other properties (such as annotations or the number of times visited). HistoryGraph also provides the means to create sets of nodes. either automatically via pattern matching on the tree, or when a notification is received. Once a set is created, it may be sent to another desktop application for processing. Likewise, individual sets and entire trees can be saved and retrieved from local storage or via the Web: assigning the MIME type (*application/X-historygraph*) to HistoryGraph trees permits these trees to be automatically loaded via a browser for viewing and manipulation.

HistoryGraph is able to receive notifications from other desktop and group applications and use these notifications to update the tree display, change the icons for documents, and update sets and properties. One group service that sends notifications is the Mediator, to indicate the change of state for documents it manages.

3.2. Group Services

The Mediator prototype incorporates two group services: the Web Page Control system and the Team Management system. The main purpose of a Mediated shared Web is to make it easy for individuals and services to discover what documents are available and to easily track and synchronize document changes and group activities.

The idea of a Mediated shared Web extends beyond simple HTML documents and encompasses two additional notions. The first is that the document space may include more active content such as CGI programs, Java code, or Web activity trees, and that these resources too may have states, owners and access controls. The second notion is that information about the members of the team can be stored in a central repository: this allows greater coupling between the team's documents and members (such as associating document permissions and classes of notification with individual members or subgroups of the team) as well as allowing this information to be accessed by multiple applications.

The Web Page Control System is used to control access to Web documents, maintain versions of these documents, and manage their state information. Mediated documents are pages that have owners and controlled access. The Mediated access service provides an index of documents as well as producing the original documents to authorized users when requested. This use of the Mediator thus creates a "shared Web": each shared document may be modified by one team member at a time, until the document is sealed, at which point it may no longer be modified. The person who is currently authorized to modify the document is known as the delegate.

The shared document space has a structure associated with it, including a project page which organizes documents by their relevant categories, a team page which lists team members, and a document index which lists documents, their states, and associated projects. In practice, we have found the document index most useful.

The Team Management System provides a directory of information about team members as well as a means to associate document permissions with members. We call our prototype of this system "The User Profile service": it is accessed by a number of group services, including a document annotation service [Schickler 96].

3.3. Notification Framework

The Notification System is used to multicast event notifications which can be received by various programs. Most commonly these programs simply display information about a notification to an individual user, but they can also take more sophisticated action, including sending out further notifications. Within the context of our Intranet workspace, these notifications are used to inform team members of the creation or deletion of documents, as well as document content and status changes. Notifications are sent to location-independent names, may be stored on behalf of disconnected users, and may be ignored by users who decide not to subscribe to particular categories of notifications. Our prototype of this Notification System [Meeks 97] provides two types of notification services. The first is a notification package that works by linking with the Zephyr Notification System [Dellafera 87], developed as part of MIT Project Athena. Zephyr provides a notification system that is location-independent (messages are sent to the name and the system deals with finding the recipient), and subscription-based (recipients may choose whether or not to receive notifications of various types). Our Notification System provides a high level abstraction layer for automatically subscribing and unsubscribing to notices, for handling notices as they are received, and for synchronizing responses to notices.

The second service is a notification recording service that is built on top of a network blackboard. The network blackboard is simply a shared information space that can be accessed using the notification package and a standard set of functions to read and update the information space. Because the blackboard itself uses the Notification System, its location is transparent to other applications that desire to use its services. In addition, the blackboard can act as a surrogate for another application that is temporarily disconnected from the network, storing notices for later delivery when the application requests them. The notification recording service does this by having the blackboard subscribe to the same set of notices that the application wants to receive.

Using the notification package and recording service allows desktop and group applications to send notifications of various kinds easily, these notifications then may be easily received and displayed by multiple users and services which subscribe to those notifications. Entities that occasionally disconnect from the Intranet can continue to participate in the notification process by having their notices saved for later retrieval.

Combining desktop applications with group services via a flexible notification service permits new combinations to be created, including novel applications which were not planned when the original services were created. For example, our HistoryGraph was easily modified to display information about mediated pages, even though HistoryGraph was initially implemented without such a notion. Our notification facility also allows new applications, such as group annotation, to participate in the "Web of services" without requiring a major architectural change.

4. Other Approaches

Nelson [Nelson 96] describes the design of a complex hypermedia-based engineering information management system. Their system is much larger in terms of its requirements for scalability and complexity; nevertheless, their system is quite similar to the one that we have developed. Of particular interest is the stage where the engineer notifies management of completion of a portion of the design; the manager can indicate approval by "signing" the document by using a digital signature. Our system could provide such a facility by using the "authenticated notice" mode supported by the underlying Zephyr system.

Bentley [Bentley 97] describes the use of events in the Basic Support for Cooperative Work (BSCW) project. Events in BSCW are stored by the system and presented to each user as part of the workspace: the user can explicitly "catch up" on certain events. A complete event history is kept: this contrasts with our approach where 1) notices are ephemeral by nature, and 2) only the last state of the document is kept (although the Mediator supports the listing of a complete revision history of the document). Finally, their proposed use of email notification based on user interests is a step towards our subscription-based design philosophy.

5. Conclusion

The next stage beyond writable Webs is the ability to support sophisticated collaborative applications. Advancing to this stage requires two major improvements. The first will be enhanced Web services that support notification of change to underlying Web objects, the management and query of meta-data concerning the various objects managed by a given Web service, and the application specific modification of documents as they are stored and fetched from the server. The second improvement will consist of desktop tools that enhance browsing activity and collaboration: these tools will offer close integration with existing browsers and can leverage the browser's ability to both fetch and display content.

We believe that powerful subscription-based notification systems can integrate the desktop tools with these enhanced Web servers, and can be used to quickly generate new application services that enhance the browsing experience, the generation of content, and the management of project activities. Such capabilities already exist: [Bowles 97] describes the use of notification (events) in the context of mission-critical applications. We believe that experimental systems such as described above both illustrate the need for such systems as well as illuminate the capabilities and directions that such systems will ultimately take.

6. References

[Apache 97] *Publishing Pages with PUT.* http://www.apacheweek.com/features/put

[Bentley 97] Bentley, R., Appelt, W., Busbach, U., Hinrichs, E., Kerr, D., Sikkel, K., Trevor, J., Woetzel, G.(1997). Basic Support for Co-operative Work. *International Journal of Human Computer Studies: Special Issue on Novel Applications of the WWW*, Spring 1997, Academic Press. http://bscw.gmd.de/Papers/IJHCS/IJHCS.html

[Bowles 97] Bowles, M. E. (1997). Publish/Subscribe in Mission-Critical Applications. *Distributed Object Computing*, 1 (5), ONDOC, Framingham, MA, 44-47.

[Brooks 95] Brooks, C., Meeks, W. S., Mazer, M. S. (1995). An Architecture for Supporting Quasi-agent Entities in the WWW. *Intelligent Agents Workshop Proceedings: Conference on Information and Knowledge Management*, http://www.osf.org/RI/www/waiba/papers/CIKM/CIKM.html

[Composer 97] *Netscape Composer: Fully Integrated, Web Document-Authoring Tool.* http://www.netscape.com/comprod/products/navigator/gold/index.html

[Dellafera 87] Dellafera, C. A. (1987). *The Zephyr Notification System*. MIT Project Athena documents, Massachusetts Institute of Technology, Cambridge, MA, USA, 1987.

[Frontpage 97] *Microsoft FrontPage Home Page.* http://www.microsoft.com/frontpage

[Hirsch97] Hirsch, F.J., Meeks, W. S., Brooks, C. (1997). Creating Custom Graphical Web Views Based on User Browsing History. *Poster Session proceedings: Sixth International World Wide Web Conference.* http://www.osf.org/www/waiba/papers/www6/webhist.html

[Mediator 96] *Distributed Authoring with the Mediator*, http://www.osf.org/RI/PubProjPgs/med-one.html

[Meeks 97] Meeks, W. S., Brooks, C., Hirsch, F. J.(1997), Staying in the Loop: Multicast Asynchronous Notification for Intranet Webs. *Proceedings of the first Australian World Wide Web Technical Conference*, 115-127. http://www.osf.org/www/waiba/papers/aw3tc/notif.html

[Nelson 96] Nelson, P., Poltrock, S. E., & Schuler, D. (1996). Industrial Strength Hypermedia: Managing Engineering Information with Hypermedia. *SIGOIS Bulletin*, 17(2), ACM Press, 18-33.

[Schickler 96] Schickler, M., Mazer, M., Brooks, C. (1996) Pan-Browser Support for Annotations and Other Meta-Information on the World Wide Web. *Computer Networks and ISDN Systems*, 28(1996). 1063-1074, http://www5conf.inria.fr/fich_html/papers/P15/Overview.html

Acknowledgments

This research was sponsored in part by the Defense Advanced Research Projects Agency (DARPA) under the contract number F19628-95-C-0042. The views and conclusion contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Defense Advanced Research Project Agency or the U.S. Government.

We also wish to thank our colleagues Dr. Murray Mazer, Matt Schickler, and Doug MacEachern for their contributions to this research endeavor.

Developing and Integrating a Web-based Quiz into the Curriculum

Angela Carbone
Department of Computer Science
Monash University
Australia
Angela.Carbone@cs.monash.edu.au

Peter Schendzielorz
Department of Computer Science
Monash University
Australia
cadal@ozemail.com.au

Abstract: In 1996 the Department of Computer Science, Monash University, implemented a First Year Advanced Students' Project Scheme aimed at extending and stimulating its best first year students. The aim of the scheme was to give students the opportunity to work on a project that best suited their needs and captured their interests.

One of the projects which became known as CADAL Quiz, (Computer Aided Dynamic Assessment & Learning Quiz), involved designing and implementing a World Wide Web (WWW) based multiple choice quiz generator and assessment tool.

Unexpectedly, at the time, there were several academics wishing to move away from the traditional mode of educational assessment and move towards interactive, computerised assessment. As a result, CADAL Quiz was incorporated into the First Year Computer Programming unit and utilised by lecturers, tutors and students.

This paper reports on a pilot project for developing and integrating CADAL Quiz into the curriculum. It highlights the unique quiz features, and its use by students and staff. The paper describes how the quiz was incorporated into the First Year Computer Programming unit and presents a conduit of attitudes useful to those who are planning to use the Web as a resource for educational assessment.

1. Introduction

With the onset of the Internet, in particular the World Wide Web (WWW), it has been increasingly popular to move away from the traditional mode of education and move towards a more interactive, computerised system. [Godfrey, 1996], [Conway, 1993], [Conway, 1994]. Such a move was simplified by the work of a devoted student who designed and implemented CADAL Quiz as part of the 1996 Computer Science, First Year Advanced Students' Project Scheme [Carbone, 1996], [Carbone, 1996].

CADAL Quiz is a multiple choice quiz generator and assessment tool that utilises the WWW. In 1997 CADAL Quiz was incorporated into the curriculum by first year lecturers. This change from paper based assessment to WWW based assessment is described in this paper as well as the various features of the CADAL Quiz package.

2. Design and Description of CADAL Quiz

CADAL Quiz, as used by the First Year Computer Programming unit, was designed with a number of goals in mind, including:

1. *Providing shared ownership of assessment questions by tutors and lecturers*. In the past lecturers took the sole responsibility of setting assessment tasks. CADAL Quiz was introduced to provide a structure for tutors and lecturers to work as a team and share in the responsibility of developing assessment questions.
2. *To encourage metacognitive learning in students*. One way of making students more aware of their learning is to provide them with self assessment questions that are tied to each week's laboratory tasks enabling them to monitor their own understanding. Unlike ordinary paper quizzes, both the student and staff can gain immediate feedback on their understanding and results. If a student is unsure of a response or result, they can discuss it with relevant staff immediately, rather than waiting for a paper quiz to be marked and returned, when the query may be less relevant.
3. *Reducing the opportunity to copy or cheat in tests.* As each quiz is unique, it makes it difficult for students to cheat.
4. *Cutting the cost of assessment.* CADAL Quiz automatically generates and corrects quizzes. This reduces staff hours required for printing, administering and marking.
5. *Recording student results.* A complete log is kept on who attempted the quiz, when they attempted it, specific question choices, time taken and final result. This not only serves as an assessment tool, but can be utilised in future curriculum development.
6. *Flexibility for staff to govern the test.* Staff decide on the number of questions per quiz, specify time slots and passwords to restrict quiz access, have the ability to view quiz logs and graph statistics.

While there are tools to develop interactive lessons on the web, such as SAMaker [Sloane and Dyreson 1996], and others [John Tasker, 1997], [Indiana University, 1997] many do not generate random sets of questions and do not record student results. The quizzes that are also online are either rigid or hard coded. These types of quizzes have limited applications and are generally useful for student self-assessment only. In overcoming the restrictions of conventional paper quizzes, a number of features were incorporated into CADAL Quiz, including:

2.1 Random ordering of questions and the A, B, C, D choices.

Quizzes are generated from a database of questions. This involves randomly selecting a specified number of questions from the database, and randomly ordering the chosen questions. The multiple choice alternatives (A, B, C, D) are also randomly ordered, so that if two students have the same question, the questions will appear to be different, which minimises cheating.

2.2 Results handling and analysis features

a) Immediate assessment with logged details.

The fact that each quiz is corrected immediately is a major advantage over paper quizzes. At the same time, student results are logged and can be analysed immediately. Logged information includes the student's name, ID number, email address, demonstrator's email address, the quiz attempted, the time it was attempted, the time taken to complete the quiz, the student's response to each question and the final result.

b) Results optionally displayed to the student.

During self-assessment, it is acceptable for the student to see their results immediately. However, in the case of a test, results can be hidden from the student until after all students have attempted the test, which again aims to minimise cheating. The same applies to emailing students their results for future reference.

c) Results optionally emailed to staff/supervisors.

In the case when the quiz is used as part of laboratory preparation, it is convenient to have the student's results emailed to the lab demonstrator for recording. This also applies if the quiz is used as a survey. This can be turned off if not required.

d) Results and statistics can be viewed and graphed online.

Logged information can all be viewed online. In the case of student responses to questions and final results, these can be graphed online, to indicate the more difficult/easier questions. This helps locate student strengths and weaknesses. The fact that this can all be done straight after or during the administration of a quiz, means that a difficult topic can be revised as soon as it becomes a problem.

2.3 The ability to subdue the randomness and specify a question breakdown.

To ensure that certain questions or topic areas are included in the quiz a question breakdown can be specified. Questions can be chosen from a range of questions, for example, select 3 questions from questions 1 to 10 or include question 24. This then guarantees that ranges of questions are selected, perhaps ensuring that several harder questions are included.

2.4 Administration features (such as adding and viewing questions online).

To make it easier for staff to insert questions into the quiz databases, questions can be added online. This includes password restricted access and step by step instructions to adding a new question. Staff can also view all questions in the database online, as opposed to a student who only sees a random portion.

2.5 Restricted access to quizzes and secure staff areas.

Most quizzes are available at any time, but in the case of tests it might be necessary to restrict access to certain people at various times. For example a test can be conducted over the course of a week, and only certain people can access it at any one time using time specific passwords. Access to staff areas is also password restricted.

3. Integrating CADAL Quiz into First Year Computer Programming

In the past it has been common practice to assess First Year Computer Science students through laboratory exercises, a multiple choice mid-semester test and an exam. The laboratory exercises are marked out of 10; 3 marks towards preparation and 7 marks devoted to the programming exercises. The mid-semester test was conducted during the lecture and counts towards 10% of the student's overall result for the subject [Farr and Nicholson, 1996].

With the traditional practices of assessment there has been concern about students copying preparation work and whether the mid semester test was cost-effective given the current pressure on resources and budget cuts. This year, CADAL Quiz changed the way in which students were assessed.

Each week a number of tutors devised and submitted a set of multiple choice questions into the database. These questions were related to the current week's laboratory task and aimed at testing whether the students had adequately prepared for their laboratory tasks and understood the abstractions of the lesson.

Students generated and attempted CADAL Quizzes during three practical classes. These quizzes contained 10 questions, chosen from a much larger set, and contributed to the preparation component of the student's practical mark for that week. In general the quiz took on average, approximately 10-15 minutes in most classes to administer and complete and in that time the students and tutors received details of the student's attempt. Some of the details shown in Table 1 below included: the date when the quiz was taken, the questions answered and a score out of 10. These results were automatically mailed to the demonstrator and counted towards 3 preparation marks.

```
Date:             Mon, 14 Apr 1997 14:09:35 +1000 (EST)
From:             Online CSC1011 Quiz
Subject:          Student Results - CSC1011 Quiz
Reply-To:         username@student.monash.edu.au
Supposedly-From: John Smith

Student:          John Smith
ID:               12345678
Demonstrator:     tutor@cs.monash.edu.au Quiz: csclab3
Date & Time:      Mon Apr 14 14:09:34 1997

Results: (3=B) (11=C) 4 (17=B) 21 (29) 1 (18=A) 24 2
** TOTAL:  5 out of 10 **

Key: X is correct, (X=Y) is incorrect, (X) is not answered.
(Where X is the question number and Y their incorrect choice.)
```

Table 1: Sample email sent to staff and students

During week 5 the students were familiar with the operations of CADAL Quiz so the lecturers used CADAL to replace the traditional mid-semester test, (which was formerly run in the lecture theatre over two lectures). The mid-semester test contained 50 questions chosen from approximately 165 questions and was held in the computer labs over a period of one week. The randomness of the questions was subdued so that tests of comparable difficulty were generated.

4. Responses - Statistics and the Educational Impact

During its first semester of operation, the impact of incorporating CADAL Quiz was measured by gathering student performance statistics and perceptions from tutors and lecturers. From the first trial there have been beneficial effects for the tutors and course lecturers as well as students.

4.1 Tutor and Demonstrator Responses

Teaching staff were surveyed to provide feedback on the effect CADAL Quiz had on the operations of the laboratory classes. They were also interviewed to discuss the feasibility of designing and shaping educational assessment tasks in groups, with combined tutor and lecturer involvement.

Although the process of formulating questions and adding them to the general pool via the Web increased the sense of shared-ownership felt by the tutors there were several deficiencies in structure and execution of the quiz identified by the group. These included:

errors in the wording of submitted question,
students could make multiple submissions,
the random ordering of questions did not ensure that all quizzes were of an equal level of difficulty,
difficulty in helping students when they answered a question incorrectly because it was difficult to tell which question they answered due to the random ordering

As a result of the above observations, the design of CADAL Quiz was changed prior to the mid semester test to provide focused, more personalised assistance to the students. The significant changes were:

- test questions were attempted and proof read by three independent tutors for better monitoring,
- students were only allowed to make one submission,
- addition of the ability to subdue the random generation of questions to produce tests of equal difficulty,
- releasing the total database of questions and answers, after the test was completed by all students, so that students could tell which question they answered, and the option they selected

The majority of tutors (70%) believed that the quiz was an effective way of determining whether a student had adequately prepared for the lab. Errors, both system and question design were rarely encountered.

4.2 Lecturer Review

The above changes in the structure and execution of the CADAL Quiz appear to have been very successful. Indeed feedback from lecturers was very positive under the revised framework.

CADAL Quiz automatically compiles and graphically displays the alternatives students selected for each quiz. The lecturers found this information very interesting and the online graph of the overall performance on each question very appealing. Not only has this enabled easy detection of the hardest questions (ie. most wrong answers) but even which wrong answer is most commonly selected. As a result, lecturers have received hitherto unknowable feedback about the meanings their students are constructing.

"It really did help me pick up quickly on where the strong and weak points are.... It was very good to be able to see, at a glance, which questions they were very good at, which questions they were on average completely clueless about (4 bars of roughly similar length), which questions they had some vague idea about but were thwarted when it comes to detail(perhaps a couple of good sized bars, other small bars), and which (few) questions completely threw them" Graham Farr, CSC1011 Lecturer

The graph produced from student results determined CADAL Quiz's usefulness in steering course design. Difficult and easy questions were highlighted so that course lecturers could accurately locate the most misunderstood topics, or poorly worded questions and answers. This has allowed improvements to teaching while the course is still running.

"I will be looking in future lectures to further emphasise some of the many points where they are weak.." Graham Farr, CSC1011 Lecturer

4.3 Student Results

CADAL Quiz was particularly useful in making students more aware of their own learning. In particular students decided whether they needed to do one or more of the randomly generated quizzes. These thought processes are all associated with enhanced metacognition.

With respect to the mid-semester test, a total of 365 students completed the test of 50 questions, with an average result of 60% (Standard Deviation 7.55, Range 7 - 49), which is comparable to the 1996 mid-semester (traditional hardcopy test) result, where 395 students sat the test and received an average mark of 60% (Standard Deviation 7.53, Range 7 - 49). With CADAL Quiz there was no indication that the students who completed the test later in the week had an advantage. The percentage of correct answers for each question varied from 98.5% (1 incorrect response out of 67) indicating a particularly easy question, right down to 10% correct and below, indicating particularly hard or poorly worded questions.

5. Conclusion and Future directions

CADAL Quiz is an application of WWW technology that has had a significant impact on educational assessment materials on the Internet. The process of formulating questions has increased the sense of shared ownership felt by tutors for the course. The lecturers have received hitherto unknowable feedback (from the aggregate statistics) about student's understanding and the meanings their students are constructing. This has allowed lecturers to adjust their teaching while the course is still running.

The quizzes were particularly useful in testing the students understanding. The personalised feedback made the students more aware of their own learning, hence enhancing their metacognitive skills.

CADAL Quiz has several advantages over paper based quizzes. These include ease of automatic marking, ease of creation of individualised tests, and immediate feedback to students. Continued changes and improvements will make CADAL Quiz one of the most functional Web-based testing methods available.

6. Acknowledgments

The author wishes to thank the first year Computer Science lecturers Dr Graham Farr, Dr Ann Nicholson, for using CADAL Quiz in their course and all the tutors for their dedicated efforts in devising questions that undoubtedly contributed to the success of CADAL Quiz.

7. References

[Godfrey, 1996] Godfrey, R. (1996). The Word Wide Web: A replacement, displacement, supplement or adjunct of traditional methods?, *Thirteenth Annual Conference of the Australian Society for Computers in Learning in Tertiary Education*, Australia, pp 221-234

[Conway, 1993] Conway, D. (1993). Improving Educational Outcomes with Computer-Mediated Presentation, *International Conference on Computers in Education (ICCE'93)*, Taiwan, pp.325-327

[Conway, 1994] Conway, D. (1994). Student Responses to Hypermedia In Lecture Theatre: A case study. *World Conference on Educational Hypermedia (ED-MEDIA'94)*, Canada, pp 141-146

[Carbone, 1997] Carbone, A. (1997). Retaining the Best Students - A First Year Advanced Students' Project Scheme. *Second Annual Australasian Conference on Computer Science Education, (ACSE'97)*, Australia, pp 141-148

[Carbone, 1997] Carbone, A. (1996). The Evolution of a First Year Advanced Students' Project Scheme , *Technical Report 97/314*, Department of Computer Science, Monash University, Australia.

[Sloane and Dyreson, 1996] Sloane, A.M. & Dyreson, C. (1996). An Interactive Self-Paced Learning Environment for the World Wide Web, *First Australasian Conference on Computer Science Education* , Australia, pp.344-351

[John Tasker, 1997] John Taskers Quiz Quest (1997). http://dialspace.dial.pipex.com/town/plaza/fg50/

[Indiana University, 1997] http://www.best.indiana.edu/quizsite/quickstart.html (1997). QuizSite Quick Start - Bureau of Evaluative Studies & Testing, Indiana University

[Farr and Nicholson, 1997] Nicholson, A. & Farr, G. (1997). CSC1011 Computer programming, *Lecture Notes*, Monash University, Australia

Putting Usability First in the Design of Web Sites

Jim Carter
University of Saskatchewan, Saskatoon, SK, CANADA
carter@cs.usask.ca

Abstract: The increasingly large variety of Web sites gives most users considerable choice in selecting which sites to use. The usability of a given site provides a major criteria in this choice. *Usability First* provides an approach based on a combination of usability testing and task analysis methods that can easily be used throughout the development of Web sites to improve their usability and thus their desirability to potential users.

The Need for a Better Approach to Design

The World Wide Web has experienced phenomenal growth due to its combination of ease of use (via browsers such as Netscape and Internet Explorer) and ease of development (via HTML and various authoring tools). This high level of usability has even enticed many of its users to become developers. However, the results of well meaning developments by both experienced software developers and users (come new to development) have often been very disappointing. Experienced developers may have difficulties in adapting their experience to the new media and culture of the Web. Users who appreciate the potential of the Web may fall short of achieving their goals due to their lack of training in systems development. Both of these groups need an easily usable approach to successful Web development. We can put *Usability First* by taking an approach based on HCI principles and techniques. This approach combines the interests of both users and developers by showing how developers can easily design highly successful web sites based on a thorough analysis of the "needs and characteristics" of the users.

"Cruising the Web" one easily encounters the good, the bad, the ugly, and often the absurd both in content and presentation. A quick survey of the less appealingly designed sites suggests the following (faulty) design principles may be at work a little too often:

minimalist design - just focusing on putting the content on-line
maximalist design - using as many bells and whistles as possible, whether they're needed or not
mimicry design - copying the design of a number of other sites, whether or not it applies
ego filled design - based on an individual's personality, with little regard for the feelings of others
paper based design - copying the design of printed documents into the computer
patch work design - where different pages each have different and often conflicting designs

Note that many of these "principles" are paired off with another "principle" which is more or less opposite to it. This suggests that there are all kinds of directions in which design can go wrong. The common factor with each of these heretic design principles is that they are focused on the developer rather than the user. They are often chosen to reduce the effort put into the development rather than the effort required by the user to use the resulting software. In short, they miss recognizing the importance of the usability of the software. They may even go so far as to assume that what they are providing is so important to the user, that the user should be willing to struggle to get it. When it comes to web sites, few people are willing to struggle. If the site isn't usable, web surfers will just pass it by for another.

A far more suitable set of design principles, which summarize the need for good human-computer interaction, are found in Part 10 of ISO Standard 9241 [ISO 96a] {suitability for the task; self-descriptiveness; controllability; conformity with user expectations; error tolerance; suitability for individualization; suitability for learning}. Each of these principles improves the usability of the software by directing the developer to concentrate on the needs of the user.

Identifying vs. Designing Usability

What makes for usability? Unfortunately this is one of the most difficult questions in the field of design. It is far easier to identify factors that contribute to a lack of usability than those which will ensure the presence of usability. Many sets of guidelines and style guides focus on specific design components. While they help designers to avoid major disasters, something extra is still required to help designers produce a highly usable system. This problem is in large measure due to the fact that usability involves the combined functioning of all components of a system. Usability is high when all components work well together producing the extra benefits of their synergy. Synergy benefits are possible even with less than optimal components. This is not to suggest that guidelines for individual components are not important. However, to achieve usability the designer must go beyond the design of these individual components and include the design of how these components interact. Current guidelines and style guides seldom provide thorough guidance on this higher level of design. Where it is provided, as in ISO 9241 Part 10, this guidance is often too far removed (by being so general) from its application to individual components or even to groups of components.

Part 11 of ISO Standard 9241 (Usability) [ISO 96b] provides a framework for specifying usability involving:
  Context of Use {users, equipment, environments, goals, tasks}
  Usability measures {effectiveness, efficiency, satisfaction}
  Specification and evaluation of usability during design

Even if we can specify usability, the question remains how to design for it. Often products are designed and constructed first, and then subjected to token usability testing just prior to delivery. At this late phase in the development cycle, little can easily be done to improve usability. Changes to the product are often only made when testing uncovers a catastrophic flaw. The track record of many years of neglecting usability suggests that it's time to put *Usability First*. The realities of the Web (where most users access most sites voluntarily and are free to ignore those sites that don't meet their needs) further indicate that if we fail to put Usability First, we are probably wasting the rest of our development efforts. What we need is a method to help us put Usability First.

Increasingly complex life cycles are being defined (within software engineering and related fields) in order to "better capture" the essentials of software development. While the increasing number of processes being defined often include additional testing processes, little additional attention is provided to the actual usability of the resulting systems. ISO/IEC JTC1/SC7 is currently developing an international standard for software life cycle processes [ISO/IEC 93] (from a software engineering perspective). It defines five primary life-cycle processes {Acquisition, Supply, Development, Operation, and Maintenance}, nine supporting life-cycle processes and four general life-cycle processes. Software development in organizations often involves a number of additional processes beyond those directly identified in the software life cycle. The Software Process Improvement and Capability Determination [ISO/IEC 94] standard identifies five groups of processes that are important to the development of software in organizations from a quality assurance perspective {customer-supplier processes, engineering processes, project processes, support processes, and organization processes}.

Neither of these approaches (nor any other of the major software engineering or quality assurance approaches) deals specifically with the unique needs of the user in terms of human-computer interaction, despite the growing proportion a typical system that involves the user interface. These approaches generally assume that if the system meets the data/information processing needs of the user, that it will be usable by the user. These types of approaches place more emphasis on documenting systems and training users to use them than on developing easy to use systems that require little or no documentation or training.

The increasing complexity of these life cycles leads to additional problems with their own usability. Most methodologies are highly technical and are designed by experts to be used by experts or at least by highly trained professionals. Only a small minority of those people who are now attempting to develop Web sites have the background required to try to use them. Furthermore, most methodologies (with some notable exceptions such as the Object Modeling Technique [Rumbaugh 91]) expect to be followed correctly and completely. However various studies [Roson 88, Glass 95] have found that even developers who think they are following a methodology often do not follow it completely, if at all.

Introducing the *Usability First* Concept

It is relatively easy to identify usability problems - especially if you're a user (rather than the developer who's ego is all wrapped up in a particular design). While much has been written about formal approaches to usability testing we all are capable of designing and conducting usability tests without the necessity of referring to these noble tomes. While the usability tests we might design may be less complete than more formal procedures, they are far better than not performing the more formal ones due to lack of tools or expertise. In fact, every day we probably perform many informal usability tests of various products or systems that we have encountered for the first time (or in a unique set of circumstances). Life, after all, includes a never ending series of new experiences. The question we need to concern ourselves with, is what do we do with the results of these usability encounters? There is a wide range of answers to this question:

- act unconsciously as if nothing new has happened
- run away from the challenge posed by the new situation
- do the minimum to succeed (often with a maximum of grumbling)
- learn from the experience (how to overcome its challenges and those of similar others)
- design a better way of doing the task at the center of the experience

We often are satisfied at being able to learn how to use the existing tools to accomplish a given task and leave off the role of design to "professional designers". Professional designers combine a thorough understanding of the tools at their disposal with a newly acquired understanding of our actual needs (especially in terms of usability). While end users may know less about available tools, they generally know much more about actual needs and can easily learn to specify tools similar to those they have experienced elsewhere (such as on other Web sites).

The basic concept behind *Usability First*, is that *a thorough consideration of usability issues belongs at the start of each phase of the development life cycle, even the initial analysis phase.*

This does not require following a single formalized and prescriptive approach to usability. Instead, the approach to usability should be one that is appropriate to both the people involved and the system being developed.

Usability First is not a fully developed methodology. It is an approach that can be used with other approaches to improve (and even to simplify) the development process. It is an attitude that can be stated simply and applied broadly. *Usability First* involves continual evaluations throughout the life cycle that include:

- evaluating the usability of methods and methodologies for developers
- evaluating the usability of applications, designs and developed systems for users

It can be used for selecting and modifying methodologies for the benefit of both developers and end users. It applies usability testing and improvement as a major process activity throughout the development of a product.

Usability First goes beyond a mere concern for the user and concerns itself with all facets of usability both for the user and for developers. While previous approaches such as User Centered System Design [Norman 86] have focused on the importance of the user, their usability has come into question [Monk 96]. Part of the problem is that User Centered Design has often been expressed as a goal rather than an objective. Like all goals, User Centered Design can seldom, if ever, be fully achieved. *Usability First* is objective based in its belief that development decisions should be based on usability evaluations. These usability evaluations provide qualitative and quantitative information that can guide the development process.

In order to put Usability First, designers need to acquire the typical experiences of a user. The Web gives a good opportunity to deal with the usability of all kinds of information and design as a user of various Web sites. Sites to be explored should include:

potentially competitive sites (with the same or similar applications, tasks, or content) which provide basic levels of expected usability and functionality that the site being developed should try to surpass;

exemplary sites (such as those selected as: "the Best of the Web", "hot pics", "what's cool", etc.) which contain features with good usability that may be emulated or modified for use in site being developed;

examples of poor sites (such as those selected as "Worst of the Web") which contain features with poor usability that should be avoided in the site being developed.

Using Task Analysis to Identify Usability Concerns

A modified task analysis method can provide a usable framework for the exploration of existing sites and a basis for the design of the new site. This task analysis will investigate the most usable way of providing for the needs of the various potential users. "Task Analysis" is a process more commonly associated with human factors / ergonomic approaches (such as those used by specialists in human-computer interaction) than with software engineering / traditional software analysis and design approaches. Rubenstein and Hersh [84] advocated extending the use of task analyses (from the early analysis stage of development where they are most commonly used) to the conscious development ("use modeling") of how systems being designed could be used. This use modeling of proposed systems, allows designers to put *Usability First.*

The Multi-Oriented Task Analysis (MOST) [Carter 91a] methodology further expanded the use task analysis as the organizing principle behind the development of intelligent, interactive systems. MOST identified four main foci to consider:

Tasks - are specific accomplishments of a person (or group of persons). The degree of accomplishment of a task is generally more important than the method of achieving it, allowing users a selection of methods.

Applications - packages or Web sites often group a selection of tools to serve a number of tasks.

Tools - are any of the many things (computerized or noncomputerized) that help a person accomplish some task (or set of tasks). Different tools (or sets of tools) can be used to accomplish the same task. Tools exist at (and are designed for) various levels: from entire Web sites down to individual links on a page.

Users - are not all the same. Severe usability problems can occur in systems designed for a "generic" user who seldom exists.

Data - is the raw material processed by computer systems. Data can be presented in a variety of formats and can be processed to higher levels such as information and knowledge. Data provides the content for applications and Web sites.

Each of the tasks, tools, users, and data can pose their own usability concerns. Further usability concerns arise in the interactions between these foci. For example, a tool that works well for one type of user on a particular task, may not work equally well for another type of user on the same task or for the same type of user on a different task. Basic usability criteria for a Web site can come from identifying the various potential types of users, tasks, tools, and data that it can be expected to bring together. While each of these four foci are important, developers can benefit from guidance in choosing the most usable one as an appropriate starting point.

Data serves the users accomplishing their desired tasks, and should be kept subservient to both users and tasks. Considerable usability problems can arise from structuring a Web site around its content rather than around how this content will be used. Unfortunately the "Field of Dreams" syndrome of "If you build it, they will come" puts the ego of the developer ahead of the needs of the potential users.

Tools, like data, serve the tasks and users. Premature focusing on tools can lead to choosing tools that are "neat" to the developer but which are impractical due to various usability problems for the user.

Users, while of penultimate importance, are only users if they use the system.

Tasks are not only the basis for individuals becoming users, but are readily analyzed by a developer exploring various web sites. This analysis of tasks should not be limited to only those tasks currently considered part of what a Web site or application should accomplish. The analysis should be expanded to include similar tasks and other potential tasks that may not be currently performed.

Identification of tasks is just the starting point for understanding what is needed. The analysis of tasks requires that we investigate the where, when, and how of these tasks in relation to their users. It is important to note situations where the same task is performed differently by different groups of users. In the past the tendency has been to force all users to a single design of use. In order to put Usability First, we need to consider how tasks are or could be done from the viewpoint of each group of users rather than that of the developer. This involves a combining of our informal style of usability testing along with our task analysis.

While considerable information can be obtained by analyzing existing Web sites, it is important to differentiate between necessary task requirements and the limitations of existing designs. Existing tools commonly used for the tasks often include a number of design limitations that have evolved over time to become expectations. Tasks should be analyzed to consider what essential limitations or requirements they place on design and how this design may be improved. Because of this interplay between tasks and tools, their analysis often proceeds together (recognizing that both can lead to opportunities for improving usability).

In order to identify their various usability requirements and concerns, each task and tool needs to be further analyzed in terms of:

- its operational details
- its requirements of users
- where it is performed
- when it is performed
- how it communicates (with others and with users)
- how it is learned
- how errors encountered during its performance are handled
- problems that it may cause

(The MOST methodology provides similar criteria for analyzing users and data.)

Each of these topics can lead to a number of further detailed questions [Carter 91b] to guide the developer in analyzing usability concerns. For example, operational details include:

- what is the purpose of the task / tool
- is it a formal or informal task / tool
- how is the task done /tool used
- what are the alternatives to it
- how flexible / adaptable is it
- are there redundancies
  - internally / externally
  - partial / total redundancy
  - and what are the costs / benefits of retaining redundancy
- what feedback does it provide
- how accessible is it
- is it sharable / concurrently usable

Going Beyond Analysis

A consideration of the purpose of the task can lead to usability concerns such as:

- Where a number of tasks can be replaced with a single generalized task, the user must be able to recognize and accept this replacement.
- Where differences in purposes exist for a single task (whether or not it is a generalized task) the user must be able to understand the effects (or lack there of) of these differences in purposes on the task.
- Similar purposes either require similar tools or, if possible, a generalized tool. The decision to combine tools into a single tool must take into account any resulting changes in the usability of the new tool for

users of the existing tools that it is to replace. Where some users may be negatively effected, there may be cause to create a separate tool (or to retain or modify an existing tool) for their use.

If similar tools are designed, their appearance and actions should be similar. Differences in appearance and actions should be directly related to the differences in their function. Thus differences should be minimal, significant, and obvious to the user.

If a single tool is designed, care needs to be taken so that the user recognizes its multiple purposes. This can be done either via the visual design of the tool, the multiple positioning of the tool within various contexts of use, or at least via training materials used to introduce the user to the tool.

Where a tool is to operate differently in different environments/states, the state in which it is operating should be obvious to the user. Additional guidance may be required to ensure that the user operates it in the manner required by the state. The same tool (including interface objects) should not have vastly different or even contradictory purposes in different environments (states) that may be used by an individual user.

While guidelines such as these could be collected from a thorough search of human-computer interaction literature, the developer would still have to determine which guidelines apply to which tasks, users, and tools and to determine how to apply them [ISO 96c]. By applying a *Usability First* approach, the most relevant usability concerns for designing the desired Web site often can be captured and evaluated from the investigation of other existing sites. Good designs will likely incorporate them and bad designs violate them.

This recognition of usability concerns and opportunities provides a good starting point for developing usable systems. Having been sensitized to such usability concerns, the developer is encouraged to continue usability testing (even if performed informally as discussed above), and to apply it to the Web site under development. The early stages of design can involve the developer evaluating use models, while later stages should involve enlisting a variety of sample users to provide independent testing of prototypes. While end user involvement is essential to user centered design, with Usability First the development also benefits from the developer's early experiences of being a user of similar systems.

Conclusion

Usability First is not intended to replace the use of more formal methods where they are useful and are usable or where they are required by system owners. In such instances it can be used with formal methods to help ensure the usability of the resulting product (which is something that no current formal method fully addresses). Likewise, it is not intended to replace the use of more formal usability evaluations, where such evaluations would otherwise be conducted. Rather, it is intended to bring an appreciation of the need and at least an informal application of usability testing into all systems development projects and especially into the development of Web sites. It is an approach that all developers, even users who have newly come to development, can easily apply to produce better, more usable Web sites.

References

[Carter 91a] Jim Carter, Combining Task Analysis with Software Engineering in a Methodology for Designing Interactive Systems, in John Karat (ed.) Taking Software Design Seriously: Practical Techniques for Human-Computer Interaction Design, Academic Press Inc., 1991, pp. 209-234.

[Carter 91b] J.A. Carter and J.P. Hancock, A Context for Designing Adaptations: The Multi-Oriented Structured Task Analysis (MOST) Methodology, SIGCHI Bulletin, 1991.

[Glass 95] Robert L. Glass, Software Creativity, Prentice Hall, Englewood Cliffs, NJ, 1995.

[ISO 96a] International Organization for Standardization, ISO Standard 9241 (Ergonomic requirements for office work with visual display terminals) - Part 10 Design Principles, 1996.

[ISO 96b] International Organization for Standardization, ISO Standard 9241 (Ergonomic requirements for office work with visual display terminals) - Part 11 Usability, 1996.

[ISO 96c] International Organization for Standardization, ISO Standard 9241 (Ergonomic requirements for office work with visual display terminals) - Part 17 Forms Fill Dialogues, 1996.

[ISO/IEC 93] ISO/IEC Joint Technical Committee 1/ Sub Committee 7 Software Engineering/ Working Group 7 Life Cycle Management, Committee Draft 12207 Information Technology: Software Life-Cycle Processes, 21 June 1993.

[ISO/IEC 94] ISO/IEC Joint Technical Committee 1/ Sub Committee 7 Software Engineering / Working Group 10 SoftwareProcess Assessment, SPICE Baseline Practices Guide, version 0.05, May 1994.

[Monk 96] Andrew Monk, User Centered Design: Quality or Quackery, Proceedings Companion for CHI'96, Association for Computing Machinery, April, 1996.

[Norman 86] Donald A. Norman and Stephen W. Draper (eds.), User Centered System Design: New Perspectives on Human-Computer Interaction, Lawrence Erlbaum Associates, Hillsdale, NJ, 1986.

[Rosson 88] Mary Beth Rosson, Susanne Maas, and Wendy A. Kellogg, The Designer as User: Building Requirements for the Design Tools from Design Practice, Communications of the ACM, 31:11, Nov. 1988, 1288-1298.

[Rubenstein and Hersh 84] R. Rubenstein and H. Hersh, The Human Factor: Designing Computer Systems for People, Digital Press, Maynard, MA, 1984.

[Rumbaugh 91] James Rumbaugh, Michael Blaha, William Premerlani, Frederick Eddy, and William Lorensen, Object-Oriented Modeling and Design, Prentice Hall, Englewood Cliffs, NJ, 1991.

WebCiao: A Website Visualization and Tracking System

Yih-Farn Chen

Eleftherios Koutsofios

Network Services Research Center
AT&T Labs - Research

180 Park Ave.
Florham Park, New Jersey

{ chen, ek }@research.att.com
This URL: *http://www.research.att.com/~ciao/webciao/webnet*

Abstract

WebCiao is a system for visualizing and tracking the structures of websites by creating, differencing, and analyzing archived website databases. The architecture of WebCiao allows users to create customized website analysis tools and present analysis results as directed graphs, database views, or HTML reports. Within a graph view, operators can be fired from any graph node to study a selected neighborhood. WebCiao has a database differencing tool that helps creators of large websites to monitor the dynamics of structural changes closely. It also helps web surfers to quickly identify new products and services from a website. An on-line demo, *Website News*, based on the WebCiao technology, has helped sharpen our focus with its daily analysis of new web contents from the internet and telecommunications industries.

---

1. Introduction

The complexity and ever-changing nature of major websites are presenting problems to both website creators and frequent visitors to those sites. For website creators, detecting structural changes and maintaining website integrity are critical before publishing new web contents outside the firewall. On the other hand, links to new contents frequently go unnoticed by visitors because they cannot locate new stuff easily.

WebCiao is a system that analyzes web pages of selected websites, stores their structure information in a database, and then allows users to query and visualize that database with graphs or HTML pages. WebCiao can also be used to analyze the structural differences of archived web databases by highlighting added, deleted, and changed pages or links.

WebCiao was created to help both surfers and maintainers of complex websites. Web surfers can use customized queries to quickly identify new products or services related to a particular topic without manually going through individual hyperlinks and pages. Website maintainers can use it to (a) visually track structural changes made by various web page authors and (b) perform global analysis to detect missing links or orphan pages before moving pages from staging machines to external servers.

The visualization component of WebCiao was used in WebGUIDE [Douglis et al. 1996] as a visual aide to the textual differencing capability of AIDE[Ball and Douglis 1996]. This paper focuses on how to combine WebCiao's query, analysis, and visualization operators to perform various website visualization and tracking tasks. It also presents a new interface, *Website News*, as an alternative to deliver updates on website changes without complex user interactions.

We shall describe the architecture of WebCiao in Section 2, the basic query and visualization capabilities in Section 3, a major application of WebCiao, *Website News*, in Section 4, and discuss related work in Section 5, followed by summary and future work in Section 6.

2. Architecture of WebCiao

Figure 1: Architecture of WebCiao

Figure 1 shows the architecture of WebCiao, which consists of three major components:

- *HTML Information Abstractor*: *hia* extracts web pages from a website and converts them into a CQL[Fowler 1994] database according to an Entity-Relationship model, which includes URL entities such as HTML and image files, relationships such as image and text hyperlinks, and their associated attributes such as URL addresses and anchor text. *hia* allows users to specify, with regular expressions or lists, what pages to include or exclude during recursive retrievals of a website's pages. It also allows users to specify the depth of recursive search, similar to that provided by WebCopy[Parada 1996], but WebCopy simply gets pages, while *hia* also converts the pages into a database.

- *Database Differencing Tool*: *diffdb* takes two versions of a database, compares page checksums and links, and creates a difference database that consists of all pages and links with tags that specify each as added, deleted, changed, or unchanged.

- *WebCiao operators*: The WebCiao system consists of a set of query and analysis operators that read and write *virtual databases*. Each virtual database consists of a subset of entities(pages) and relationships(links) retrieved from the complete database. A set of *view operators* takes any virtual database and converts it to a directed graph, a database view, or an HTML page report. Since query and analysis operators are interchangeable, a *virtual database pipeline* can be constructed to perform complex operations before the results are turned into graphs or other forms of reports. These operators can be used on command lines, in shell scripts, or invoked by WebCiao's graphical interface, or a web interface discussed in Section 4.

WebCiao inherits our years of software reverse engineering [Chen 1995a] experience in querying, analyzing, and visualizing large and complex software structures. WebCiao is an instance of Ciao[Chen, et. al 1995b], a multi-language graphical navigator for software and document repositories. Ciao has been instantiated for C, C++, Java, Ksh, HTML, and some other languages and business databases. The architecture style shown in Figure 1 applies to all languages. Except for HTML-specific tools like *hia* and operators that communicate with web browsers, the complete set of GUI, query, and analysis tools is generated automatically from a CIAO specification file less than 200 lines long.

3. Querying and Visualizing Structure Changes

WebCiao consists of several operators that can be combined on its virtual database pipeline:

- *Selection* operators: retrieve a set of entity or relationship records according to the selection criteria.

- *Closure* operator: performs reachability analysis according to the specified level of recursion.

BEST COPY AVAILABLE

- *Focus* operator: performs fan-in/fan-out analysis in the neighborhood of selected pages.
- *Database View* operators: generate database views.
- *Graph View* operators: generate graph views.
- *Visit* operator: sends requests to a web browser to retrieve corresponding pages.

Except for *View* and *Visit* operators, all operators read and write virtual databases. Additional analysis and view operators can be written to interface with the virtual database, which is simply an archive of plain text database files that can be unpacked easily.

The following two examples illustrate how CIAO operators can run on a *difference database* created for AT&T's website based on the changes from November 27, 1996 to December 2, 1996.

- Show new web pages that match the pattern '*press*':

  ```
  $ ciao_eset url '*press*' etag=added | ciao_eview url -
  name                                               kind   etag
  ================================================== ====== ==========
  http://www.att.com/press/1196/961127.csa.html      url    added
  http://www.att.com/press/1196/961125.bsb.html      url    added
  http://www.att.com/press/1196/961127.cia.html      url    added
  ```

  ciao_eset is an entity selection operator, while *ciao_eview* is an entity database view operator. The result indicates three new press releases from AT&T during that period.

- Use a graph to show changes in the neighborhood of AT&T's Easy Commerce page:

  ```
  $ ciao_focus -l2 url http://www.att.com/easycommerce | ciao_rgraph url - url -
  ```

  The focus operator *ciao_focus* studies the neighborhood of a particular URL at the specified level of depth (in this case, up to the 2nd level) and pipes the output to the graph generator, *ciao_rgraph*.

Figure 2 shows the result of the last query. Changed web pages are shown as yellow ellipses, deleted web pages as white rectangles, while green nodes represent those pages that stay the same. New pages are usually shown as red rectangles, but we don't have any in Figure 2. The picture allows us to easily identify changes in incoming and outgoing links. If the Easy Commerce page has to be modified, deleted, or moved, we know what other pages need to be checked or updated.

As an example of more complex operations, suppose we are interested in finding information under a particular node on a website, similar to the functionality provided by GlimpseHTTP[Klark and Manber 1996] (and recently, WebGlimpse[Manber et. al 1997]). In WebCiao, we can simply run a *closure* operator performing reachability analysis on the selected node followed by a *selection* operator based on the URL addresses, anchor text of each link, or page contents (if archived). For example, the following virtual database pipeline reports the set of URL's in the first three layers of pages reachable from *http://www.att.com/news* whose addresses match the pattern "**worldnet**" on December 10, 1996:

```
$ ciao_closure -13 url 'http://www.att.com/news' | ciao_eset url '*worldnet*'
| ciao_eview url -

name                                                                     kind
======================================================================== ======
http://www.att.com/w3403/attworldnetservice/crystal.html                 url
http://www.att.com/worldnet/wis/sky/signup.html                          url
http://www.att.com/worldnet                                              url
http://download.worldnetall.com/mainL54Q.htm                             url
http://www.att.com/w3403/attworldnetservice/legall.html                  url
http://www.worldnet.att.net                                              url
http://www.att.com/worldnet/wis/game/gamstrt.html                        url
http://www.att.com/worldnet/wis/                                         url
...
```

Figure 2: Changes in the Neighborhood of AT&T's Easy Commerce Web Page (November 27, 1996 to December 2, 1996)

4. Application: Website News

A demo called Website News[Chen and Koutsofios 1996a] has been set up to demonstrate applications of the WebCiao difference databases. We archive the home pages of a selected set of frequently-visited websites (such as AT&T, Microsoft, Netscape, and IBM) on a daily basis. Users can find new links added everyday on these websites by using the *Website News* web interface page shown on the lefthand side of Figure 3. The righthand side shows the news report on Monday, December 9, 1996 for the selected websites. Users can click on any of these new links directly to get to the new pages without going through the home page and figuring out the new links added. The report shows, for example, that AT&T has added a link to a page about *missing children*, Microsoft added a web page about *Job Opportunities*, etc. since last business day (Friday, December 6). A user can also review the change history of a website by picking any two dates with archived databases and make a *difference database* to see all the links added and deleted. .

In response to a request from an AT&T marketing group, we have created a separate Website News demo for the telecommunications industry group [Chen and Koutsofios 1996b], which includes the websites of AT&T, MCI, Sprint, and the seven Regional Bell Operating Companies (RBOCS). It also demonstrates that the same set of CGI scripts we developed can be reused on a collection of websites that might be of interest to a community of users. Website News could also be useful in reducing the network usage of major corporations or internet service providers by providing the change information of popular websites upfront and thus eliminating many unnecessary downloads by users.

Figure 3: Website News: Interface Page (left) and News Report (right)

5. Related Work

Recently, there have been growing interests in visualizing the complex structures of major websites -- mainly to help web users locate information faster without getting lost. Examples include WebMap[Dömel 1994], which captures a user's dynamic interactions with the web pages and visualizes the navigation history, and NetCarta's Web Mapper[NetCarta 1996](now part of Microsoft's Backoffice), which performs a static analysis of the structure of any selected website. Other examples include Web Analyzer[InContext 1996], which presents a *wavefront* view, and Hy+[Hasan et. al 1995], which is based on the visual query language GraphLog and, like WebMap, uses dynamic trace information obtained during a Mosaic session. WebCiao is similar to NetCarta as it also maps a website, but it allows users to make customized database queries and visualize changes in website structures.

Tracking website changes is critical for both website maintainers and clients. A website maintainer needs to make sure that there are no missing links or orphan pages after changes are made to a website. A frequent visitor to a website may prefer to be notified when changes occur on that website. Most website change-tracking systems or notifiers such as Smart Bookmarks[FirstFloor 1996] or AIDE[Ball and Douglis 1996] focus on textual changes, while WebCiao focuses on structure changes on a website. WebGUIDE[Douglis et al. 1996] combines AIDE and the visualization component of WebCiao, to allow the examination of both textual and structure changes in web repositories. However, the current framework of WebGUIDE does not allow global database queries and analysis operators to be performed on a set of web pages.

Website News was inspired by both WebGUIDE and Internet Archive[Kahle 1996], which has the vision of building a complete running snapshot of the public world-wide-web so that the history of anyone's favorite sites can be preserved. If the complete Internet Archive becomes a reality, our vision is that one day a user can use Website News and WebCiao not only to analyze the history of any changed websites, but to construct a search engine like AltaVista[DEC 1996] on WWW *deltas*.

6. Summary and Future Work

We have found WebCiao to be quite flexible in querying and visualizing the structures of complex websites. The difference database created by WebCiao allows us to monitor the dynamics of many major websites closely and effectively. The change information is useful in tracking evolving products and services on the web, browsing the web with limited bandwidth, and maintaining large websites. The on-line demo, *Website News*, has been serving many customers world-wide on a daily basis to deliver updates on the structure changes of several major websites. We believe that WebCiao could become extremely useful in identifying new web contents if it is applied to generate web deltas for an internet archive of public websites.

Acknowledgements

Fred Douglis wrote *html2db*, the predecessor of *hia*, in Perl, and helped with the integration of WebCiao and AIDE. Glenn Fowler wrote *cql*, the query language used extensively in WebCiao. WebCiao's visualization tools are based on *GraphViz*, written by John Ellson, Eleftherios Koutsofios and Stephen North.

References

[Ball and Douglis 1996] Thomas Ball and Fred Douglis. **An internet difference engine and its applications** . In *Proceedings of 1996 COMPCON*, February 1996, pp. 71-76.

[Chen and Koutosofios 1996a] Website News. http://www.research.att.com/~chen/web-demo, 1996.

[Chen and Koutosofios 1996b] Website News: Telecommunications Group . http://www.research.att.com/~chen/web-demo/telecom.html, 1996.

[Chen 1995a] Yih-Farn Chen, **Reverse Engineering.** Chapter 6, in "*Practical Reusable UNIX Software*," edited by Balachander Krishnamurthy, pages 177-208, John Wiley & Sons, New York, 1995. .

[Chen et. al 1995b] Yih-Farn Chen, Glenn S. Fowler, Eleftherios Koutsofios, and Ryan S. Wallach. **Ciao: A Graphical Navigator for Software and Document Repositories.** In *International Conference on Software Maintenance* , pages 66-75, 1995. See also the home page at http://www.research.att.com/~chen/ciao.

[DEC 1996] AltaVista. http://altavista.digital.com, Digital Equipment Corporation, 1996.

[Douglis et al. 1996] Fred Douglis, Tom Ball, Yih-Farn Chen, and Eleftherios Koutsofios. **WebGUIDE: Querying and Navigating Changes in Web Repositories.** In *Proceedings of the Fifth International WWW Conference* , Paris, France, April 1996. Also appears in Computer Networks and ISDN Systems, 28(1996), pages 1335-1344.

[Dömel 1994] Peter Dömel. **Webmap - a graphical hypertext navigation tool**. In *Proceedings of the Second International WWW Conference*, 1994.

[FirstFloor 1996] First Floor Software. http://www.firstfloor.com/, 1996

[Fowler 1994] Glenn S. Fowler, **cql -- A Flat File Database Query Language.** In *USENIX Winter 1994 Conference*, San Francisco, January 1994.

[Hasan et. al 1995] Masum Hasan, Dimitra Vista, Alberto Mendelzon, **Visual Web Surfing with Hy+**, Proceedings of CASCON'95, Toronto, Canada, 1995. Also visit the web page at http://www.db.toronto.edu:8020/webvis.html, 1995.

[InContext 1996] Web Analyzer. http://www.incontext.com/products/analyze.html, InContext Systems, 1996.

[Kahle 1996] Brewster Kahle, **Internet Archive.** http://www.archive.org/webarchive96.html, 1996.

[Klark and Manber 1996] Paul Klark and Udi Manber, **GlimpseHTTP** http://glimpse.cs.arizona.edu/ghttp, 1996.

[Manber et. al 1997] Udi Manber, Michael Smith, Burra Gopal, **WebGlimpse** -- Combining Browsing and Searching, Proceeding of the Usenix Conference, January, 1997. Also visit the web page at http://glimpse.cs.arizona.edu/webglimpse.

[NetCarta 1996] NetCarta. http://www.netcarta.com/, 1996.

[Parada 1996] Victor Parada, **WebCopy**. http://www.inf.utfsm.cl/~vparada/webcopy.html, 1996.

Comparative Diffusion of the Telephone and the World Wide Web: An Analysis of Rates of Adoption

Hsiang Chen
School of Information Studies
Syracuse University
Syracuse, NY, 13244, USA
hchen04@mailbox.syr.edu

Kevin Crowston
School of Information Studies
Syracuse University
Syracuse, NY, 13244, USA
crowston@syr.edu

Abstract: This paper investigates the diffusion process of the World Wide Web technology by means of a comparison with telephone diffusion at the end of the nineteenth century. The Web technology has diffused at exponentially around the world. In contrast, the telephone technology, a similar innovation of interactive communication technology imbued with typical uncertainty and impedance, took several decades to diffuse. This paper diagnoses the fundamental differences between these two innovations by analyzing their innovation perceived attributes, such as relative advantage, compatibility, complexity, trialability and observability, and attempts to explain the determinants of their rates of adoption.

Introduction

One hundred years ago, the telephone was described as "the youngest and most wonderful development of the means of communications" [Martin 1991]. In his autobiography of 1926, Watson, an important partner of Bell, stressed, "I don't believe any new invention today could stir the public so deeply as the telephone did, surfeited as we have been with the many wonderful things that have since been invented" [Watson 1926]. Today, the Web stirs the public perhaps as deeply as the telephone once did.

As a general proposition, even though the Web and the telephone are two interactive communication technologies, one new and one old, separated from each other by almost 120 years, they share some similarities. These inventions, in particular, open the doors of distributed communications, one orally and the other electronically, and allow human beings to extend their perceptions to surpass the obstacle of space. They increase the possibilities for communications and help human beings understand themselves in a more sensitive way. On the one hand, the telephone changed how we live and how we communicate. It restructured our society. The diffusion of the telephone made possible the multistory residence and office building and modern city [Brooks 1975]. On the other hand, the invention and the diffusion of Web technology created another dramatic change for human beings. It made real the concept of the global village and virtual community by creating the possibility of a universal database and the accessibility of distributed information.

Indubitably, the most significant difference between the Web and the telephone is their respective rate of adoption: the relative speed with which an innovation is adopted by individual members in a social system [Rogers 1995]. Web technology has been diffusing at an exponential growth rate, and has been establishing its bridgehead around the world in a very short period of time with little resistance. In contrast, telephone technology, a similar innovation of communication technology imbued with uncertainty and impedance, took five decades to reach 10% of the households in the United States [Fischer 1992], whereas the Web took only five years to reach the same level.

To those living in the late nineteenth century, a device to transmit actual humans voice was a completely new concept. People were scared, puzzled, and awed. Because it was invented in a relatively conservative social system, such an innovative technology took a longer time to spread out. In contrast, Web technology is compatible with its imbedded environments, builds upon the existing Internet structure, and consequent can diffuse very quickly. Of course, the fast diffusion pace of the Web innovation might be also attributed in part to the current, less conservative, social environment; but it is the nature of the innovation of the Web per se which in fact made the difference.

This paper explores and analyzes the differences of the rate of adoption between the telephone and the Web, with an ultimate aim of explaining why they differ. With the counterpoint of the telephone, by understanding their similarities and differences, this paper will then identify these determinants.

Diffusion of the Telephone

When the old technology of the telephone was new, people were in a dilemma. In the 1860s, people expected that a human being's voice could be transmitted in distance through different media but also believed that human speech was sacred and should not be carried by electricity. Thus, the very idea of the telephone generated supernatural fear and uneasiness for the public at large in the 1870s [Brooks 1975] while others thought of the telephone as a ridiculous and impractical toy. During that era, people were not able to accept the fact that a mysterious box could emit a human voice when no one was there; and this situation could only be explained by either mystical magic or insanity [Brooks 1975]. The social background and structure in the 1870s were not ripe to accept such a revolutionary technological achievement as the telephone. It is almost as if the idea of a speech-formed electric current did not cross the scientist's mind [Watson 1926].

Then, in the early days of the telephone development, it was seemingly taken to be a substitute for the Morse key, rather than a replacement for the telegraphic function itself [Garnet 1985]. In its early demonstration, even though Bell inspired awe and wonderment in public, most people remained certain that the telephone would never eclipse the widely used printing telegraph instrument [Garnet 1985]. To the business community, it seems that the telephone did not provide any tangible advantages over the existing functions of the telegraph [Garnet 1985]. People simply did not accept Bell's vision and the invention of the telephone was not seen as a threat to the telegraph by the industry itself.

In order to urge the public to accept the usage of the telephone, Bell demonstrated his "magic box" in different places to different people. In May, 1877, Bell gave his most important demonstration to Boston-area worthies. At least in Boston, the telephone had "passed out of the realm of suspected witchcraft" [Brooks 1975]. Newspaper publicity attracted people's attention, and people began to perceive the importance of the telephone [Watson 1926]. What emerged from this was that thousands of people were entirely willing to pay fifty cents to hear a lecture from Bell about how the telephone was invented and to hear how the telephone talked.

Yet, a crucial aspect was that the public had to be educated. After the initial demonstration stage, telephone salesmen inevitably had to introduce the telephone and demonstrate its utility face-to-face to potential customers. This included convincing non-English-speakers that the instrument "spoke" their languages, and that the telephone wire was not able to transmit any diseases [Fischer 1992]. For decades, most marketing experts in the telephone industry emphatically believed that to sell their product they had to find or to create uses for it. Thus, telephone entrepreneurs in the early years broadcast news, concerts, church services, weather reports, and stores' sales announcement over their lines [Fischer 1992].

In May 1877, the first experimental central exchange was opened in Boston [Brooks 1975]. Early in 1878, the usefulness of the telephone was greatly increased by the development of a workable exchange, making possible switched calls among any number of subscribers rather than merely direct connections between two or three. Late in 1879, telephone subscribers began for the first time to be designated and called by numbers rather than by their names [Brooks 1975]. The other technological advance in telephony in the 1880s was the establishment and rapid growth of long-distance service. Above all, long distance service was obviously in the public need and interest. Hence, the telephone diffusion took off and gradually linked different sectors of economic activity and became a device permeating people's daily life. By the 1920s, the telephone had reached 10% of households of the United States [Fischer 1992].

Diffusion of the Web

For several decades, human beings dreamed of the concept of a universal database of knowledge. Wells' essay on a "World Encyclopaedia" [Wells 1938] proposed the possibility of building a universally accessible archive of the entirety of human knowledge. Later, in 1945, Bush [Bush 1945] imaged a "wholly new form of encyclopedias", "with a mesh of associative trails running through them." In 1963, Weinberg [Weinberg 1963] suggested an "information transfer chain," operating like a switching system. This device would connect the user, quickly and efficiently, to the proper information and only to the proper information. Since then, organizing the knowledge of the whole world into a "world brain," [Wells 1938] and allowing everyone to retrieve from it, has been an intellectual dream in the scientific field.

Wells' "World Encyclopaedia" and Bush's "new form of encyclopedias" and Weinberg's "information transfer chain" have been realized several decades later through the implementation of the World Wide Web. Only now has the technology caught up with these dreams, making it possible to implement them on a global scale. Similar to Wells' vision of 1938, the Web was created to be a "pool of human knowledge," distributed to share human beings' ideas [Berners-Lee, et al. 1994]. Tim Berners-Lee, who might best be termed the "creator of the Web", also called this new innovation a "World Wide Brain," suggesting the analogy on grounds that "people within the Web are organized like neurons in a brain" [Berners-Lee 1997].

Berners-Lee created the technology that made the Web possible in 1990 while working for CERN. CERN is a European Particle Physics Laboratory, which is a collection of European high-energy physics researchers. The original purpose of the WWW was just to give physicists in the field of high energy the means to communicate and exchange ideas easily. He created the first World Wide Web server and the first World Wide Web client by building and combining the network protocol, HTTP (HyperText Transport Protocol), the language, HTML (Hypertext Markup Language), the address system, URI (Universal Resource Identifiers) and Internet database in the server. By the end of 1990, the first piece of Web software was introduced on a NeXT machine, designed to allow people to work together by combining their knowledge in a web of hypertext documents.

Demonstrations were given to CERN committees and seminars in 1990, and made available on the Internet at large in the summer of 1991. Later, a presentation was given at the Hypertext '91 conference. Throughout 1992 Berners-Lee continued to promote the project, as several developers began to work on their own contribution to the World Wide Web. Since then, partly due to media publicity, thousands or even millions of people throughout the world have contributed their time writing Web software and documents or telling others about the Web. That is to say in a way never envisioned by the original participants in the Web, the project has reached global proportions in a very short period of time.

Seemingly, there was a snowball effect. One gains the impression that it was very difficult in the beginning to explain the potential uses of this new information technology. Since there was little information and few Web sites available to users, the snowball at least did not roll by itself in the beginning. As an interactive medium, the Web clearly must reach its critical mass point [Markus 1987] first in order to take off. At this point, in order to get the snowball going, Berners-Lee and others did their best to push this snowball. It was considered a serious turning point for the Web diffusion when Mark Andreeson in NCSA created Mosaic, a Web client application which was available on the Internet. Mosaic pushed the snowball.

Certainly, the Web has grown rapidly. The first Web server was introduced to the world in 1991. In the beginning of 1993, there were scarcely 50 Web sites around the world. Yet, in October of that year, there were over 600 known Web servers. By June 1994, there were over 2,700 servers. The number of the Web servers doubled over less than 3 months. By June 1996, there were 230,000 Web servers, and just seven months later, the figure increased by 280% to 650,000 servers. Astonishingly, the basic Web protocol has become the primary carrier of net data traffic. While the number of Web servers increased on the Internet, more users rushed into the Internet and became clients of Web technology. One major fact emerges. It was estimated that at the end of 1996, there were approximately 45 million people using the Internet (most Internet users are Web users), with roughly 30 million of those in North America, 9 million in Europe and 6 million in Asia/Pacific. What is notable is that it has been only six years since the Web was invented.

Diffusion in Rate of Adoption

All innovations are not equivalent units [Rogers 1995]. The diffusion of the telephone and the Web varies in different ways, such as the rate of adoption, features of innovation, and relevant social system. Nonetheless, among them, the rate of adoption is a significant difference between telephone technology and Web technology. The rate of adoption means the relative speed with which an innovations adopted by members of a social system. Generally it is measured by evaluating the number of individuals who adopt a new idea in a specified period of time in a social system [Rogers 1995]. To a great extent, while the snowball effect was visible in the diffusion of the Web, it was not as apparent in the diffusion of the telephone. It can be said that the diffusion of the Web easily reached its critical mass point, which allows the Web to takeoff at a considerable accelerating rate.

The differences of the rate of adoption for innovations can be explained, according to Rogers, by perceived attributes of innovations, type of innovation-decision, communication channels, nature of the social system, and extent of the change agent's promotion efforts. Among them, the perceived attributes are the most important explanation for the rate of adoption of an innovation. About 49 to 87 percent of variance of the rate of adoption can be interpreted by the five innovation characteristics of perceived attributes*: relative advantage, compatibility, complexity, trialability, and observability* [Rogers 1995].

This section focuses on the perceived attributes of the telephone and the Web, aiming to explain the differences in the rate of adoption between these two innovations. Usually, Web technology can be divided into Web servers and Web users, and so does its rate of adoption. This study compares only the rate of adoption of Web users with the rate of adoption of the telephone. The diffusion of the potential Web servers is not discussed here.

Relative advantage

Relative Advantage can be explained as the benefits and the costs resulting from adoption of an innovation. The fact of the matter is that the degree of relative advantage is often expressed as economic profitability, social prestige, or other benefits [Rogers 1995].

1. Cost-benefit: The telephone, in its initial diffusing stage, was not a universal service, and its installation fee and usage fee were not affordable to the majority of potential users. The Web, on the contrary, while contributing additional functionality to current devices, does not increase the users' burden too much. In short, while both technologies can extend the human body's perception by reducing the obstacles in space, adopting Web technology does not cause economic disturbance in households or individuals while the telephone did. Besides, much Internet traffic is generated by academic institute, such as universities, where the Internet connection are free for students and staffs and computers are already in wide use. The adoption of the Web by the institutions incurs virtually no additional costs.

2. Preventive innovations: Usually, a preventive innovation has a slower rate of adoption because its relative advantage is more difficult to be noticed by individuals in a social system [Rogers 1995]. Adopting Web technology can obtain the advantages and benefits of the WWW for accessing distributed information immediately while adopting the telephone in the initial stage, which supported only point-to-point communication, might not be able to obtain the potential advantages, such as long distance calls or switched connection calls. For beneficial consequences, Web technology has a short time interval, and the telephone, a long one.

Compatibility

Compatibility is the degree to which an innovation is realized by individuals as consistent with the existing values, past experiences, and needs of potential adopters [Rogers 1995]. A more compatible idea in a social system is less problematic to the potential adopter and does not cause contradictory situations in the individual's life [Rogers 1995].

1. Values and Beliefs: At the time the telephone was invented, people and scientists did not believe that human speech could be transmitted through electricity. They also rejected the idea of the telephone because of a feeling that a human's voice was the gift of God and should not be transferred into electricity. The very idea of the telephone conflicted with sociocultural values and beliefs. Web technology is a different story. Before the birth of the Web, the Internet has been used widely. As things stand now, Internet users know that they can transfer files between nodes, logon to remote computers, chat in cyberspace, post items to the newsgroups, and send electronic mails. In sum, different Internet applications and protocols have been used and accepted, such as gopher, ftp and telnet. When Web technology was finally introduced, it was not much more than simply adding another new function to the field or creating an innovation by combining previous innovations. Norms were not violated at all.

2. Previously Introduced Ideas: Telegraph had been used commonly in business when the telephone was first introduced. People were deeply imbedded with the usage of the telegraph consequently and believed that the telegraph was their right communication tool. In the business community, people were depending heavily on the printing telegraph for their commercial transaction, and the telephone could not find its place in there. In contrast, Web technology did find its place in education, business and government. People were using gopher, telnet, ftp, email and other Internet functions, and expected a better tool for more efficient communications. The introduction of Web technology was not only compatible with other functions but also enhancing the functions of Internet. Even so, Web technology provided Internet users a better environment and better usage.

3. Client Needs: In the telephone's initial diffusion stage, when uses and functions of the telephone were not clear, new uses had to be created or founding in order to sell telephones. Potential customer needs were vague at that time, especially before the central exchange and long distance call were possible. Admittedly, the Web in its initial stage was different. In a very short time, users and Web masters easily located those potential power of Web technology, and the whole Web society worked on that potentiality to meet clients' fundamental needs. Unlike the telephone, the Web society did not need to "create" client needs but to meet them.

Complexity

In particular, *complexity* is the degree to which an innovation is perceived as relatively difficult to understand and use. Generally, the complexity of an innovation, viewed by individuals in a social system, is negatively related to its rate of adoption [Rogers 1995]. The telephone and Web technology do not vary much in their degree of complexity. At first, telephone and Web users might naturally feel awkward but their comfort level rapidly rises. To users' perspectives, both technologies are located at the simplicity side of the complexity-simplicity continuum.

Trialability

It may be remarked that *trialability* is the degree to which an innovation may be experimented with on a limited basis. In general, the trialability of an innovation, as perceived by members of a social system, is positively related to its rate of adoption [Rogers 1995]. For one thing, both technologies are in quite a high degree of trialability. In its early days, telephones were leased, and users did not need to buy the whole equipment. At most, Web technology is the same. Users can try using the Web by subscribing to the Internet connection from Internet Providers, or try it in schools.

Observability

Observability is the degree to which the results of an innovation are visible to others. In this respect, the observability of an innovation, as perceived by members of a social system, is positively related to its rate of adoption [Rogers 1995]. But, as both the telephone and Web technologies are hardware-oriented, embodying the technology as material or physical objects, their observability is very similar. Web technology, including computers and modems, and telephone technology, including telephone device and lines, are easily apparent to users' observation.

Conclusion

It is clear that even though both innovations are similar, the characteristics of these two innovations are different. Such a revolutionary technology as the telephone inevitably takes a longer time to diffuse throughout a relatively conservative social system. In contrast, the World Wide Web, much more compatible with its imbedded environment, has been able to diffuse with exemplary speed. Undeniably, the most important determinants of the relatively slower diffusion pace of the telephone and the faster diffusion of the Web are the characteristics of the innovations themselves and the social and cognitive environment in which they are embedded. This paper proposes that, among the five perceived attributes normally used to interpret the precise differences among adoption rates, the two most important determinants are relative advantage and compatibility.

References

[Berners-Lee 1997] Berners-Lee, Tim. (1997). World-wide computer: The Next 50 Years: Our hopes, Our Visions, Our Plans. *Communications of the ACM*, 40(2), p57-58.

[Berners-Lee, et al. 1994] Berners-Lee, T., Cailliau, R., Luotonen, A., Nielsen, H. F., & Secret, A. (1994). The World-Wide Web. *Communications of the ACM, 37*(8), 76-82.

[Brooks 1975] Brooks, J. (1975). *Telephone: The First Hundred Years.* . Harpter & Row.

[Bush 1945] Bush, V. (1945, July). As We May Think. *Atlantic Monthly, 176*(1), 101-108.

[Fischer 1992] Fischer, C. S. (1992). *America calling : A social history of the telephone to 1940.* Berkeley: University of California Press.

[Garnet 1985] Garnet, R. W. (1985). *The Telephone Enterprise: The Evolution of the Bell System's Horizontal Structure, 1876-1909.* The Johns Hopkins University Press.

[Markus 1987] Markus, m. L. (1987, October). Toward a "Critical Mass" Theory of Interactive Media: Universal Access, Interdependence and Diffusion. *Communication Research, 14*(5), 491-511.

[Martin 1991] Martin, M. (1991). *"Hello, Central?" Gender, Technology, and Culutre in the Formation of Telephone Systems.* McGill-Queen's University Press.

[Rogers 1995] Rogers, E. M. (1995). *Diffusion of Innovation, Fourth Edition.* New York: The Free Press.

[Watson 1926] Watson, T. A. (1926). *Exploring Life: The Autobiography of Thomas A. Watson. D.*. Appleton and Company.

[Weinberg 1963] Weinberg, A. (1963). Science, Government, and Information. In *Science, Government, and Information: The Responsibilities of the Technical Community and the Government in the Transfer of Information* (pp. 7-37). President's Science Advisory Council, U.S.: Government Printing Office. Washington, D.C..

[Wells 1938] Wells, H. (1938). *World Brains*. In *World Encyclopedia* (pp. 3-35). New York: Doubleday, Doran & Co., Inc.

Detecting Themes in Web Document Descriptors

David K.Y. Chiu and David R. Brooks
Dept. Computing and Information Science
University of Guelph, Guelph, Ontario, Canada

Abstract: A theme is a recurring pattern that may involve only a subset of the descriptors that describe a datum. The theme may or may not be exactly observed in the data. This paper develops a theme detection method that is applicable to a collection of data such as text descriptors in documents. The goal of theme detection is to provide a more precise and flexible interpretation of the data, thereby facilitating organization of the data if necessary. The following principles are used in selecting themes: stable cost, statistical dependency, structural dependency, and weighted towards larger theme size. An algorithm is developed that applies to web document descriptors from a search engine in real time. The results can be used to refine or reorganize the search. Preliminary experimental results demonstrate the effectiveness of the method using the AltaVista™ or Lycos™ search engines.

Introduction

Data, such as text descriptors from documents, may contain several *recurring patterns* that we call *themes*. Themes can be intuitively thought of as motifs, recurrent subjects, main ideas or basic concepts. For example, a novel has a theme if it has a central recurring idea that is expressed from time to time throughout its length. A musical composition can have a theme if a certain passage or its variation is repeated from time to time throughout the piece. A collection of paintings has a theme if a recurring image is observed in them, even though its occurrence may not be exact. Commonly in text documents, we can describe each document by a set of descriptors or key words. The set of documents has a theme if there is a recurring pattern of descriptors throughout, even though the pattern may not be exactly the same in each of the documents that observe this pattern. In general, a set of data can have several themes that may contain some common descriptors. This means that more than one theme can be observed in a document.

Normally, *theme detection* is an unsupervised learning process. It is independent to the process of classification or clustering, even though these processes may make use of the results from theme detection to generate more interpretable groupings. Theme detection is based on a subset of the descriptors rather than the complete set of descriptors of a data. It is therefore related to the processes of feature selection and clustering but not identical to them. Compared to feature selection [Kittler 1986] that is based on variables, it is based on attributes. Compared to clustering [Jain 1988] that involves all the features of the data, it may involve only a subset of the attributes. In this regard, it follows the approach of event-covering which was proposed in [Chiu 1986].

When applied to the text descriptors of extracted documents from a web search, theme detection provides an interpretation of the search results. It is an information extraction process [Lehnert 1994] that may be used to reorganize all the matching documents from a search engine. Compared to other methods of knowledge discovery in web search data (for example, [Zaiane 1995]), this approach requires much less user intervention and is more data dependent.

A Theme Detection Method

In this section a theme detection method will be described [Fig. 1]. In the next section, the method will be implemented in order to detect themes from web search data.

The following notation will be used. Let $A = \{a_0, a_1, \ldots, a_{n-1}\}$ be a set of n alphabets of descriptors (or primitives). Then, $s = s_1 s_2 \ldots s_m$ is a sequence of descriptors that comes from A, that is s_i A for $i = 1, 2, \ldots, m$. The descriptor s_i is an attribute of s. The length of s is m. Assuming that there are N samples in the data set. The set of samples can be denoted as $S = \{s^{(1)}, s^{(2)}, \ldots, s^{(N)}\}$.

In theme detection, statistical techniques are used. In addition, criteria are used for selecting the optimal *theme set* that describe the data - based on principles of: *stable cost, statistical dependency, structural dependency* and *weighted towards larger theme size*. The principle of stable cost requires that the final theme set will contain no more individual themes than required to meet the other criteria in describing the whole data set. The criterion of weighted towards larger theme size requires that a theme set with the least overlapping of descriptors is preferable, and a theme with a larger set of descriptors is preferable. This prevents a single descriptor theme that occurs very frequently from being preferred over longer and more descriptive themes that occur less frequently. The principle of statistical dependency requires that the attributes in the resultant themes be statistically dependent. This ensures that the attributes are related statistically. The principle of structural dependency ensures that the descriptors in a theme are structurally related to each other such as being observed in the same data appropriately. All these principles will be used as the basis for designing the theme detection method.

A. Attribute Selection Process

The first process of the method is attribute selection to generate a reduced set of attributes. *Attribute selection* is the process of selecting potential attributes (or descriptors) to be included in generating hypothetical themes for consideration at the later process. Procedures are used to remove attributes that are not likely to be involved in the construction of hypothetical themes. Statistical evaluation is used here. If an attribute occurs very infrequently or alternatively if it occurs almost in all the data, it is selected out. An attribute that occurs almost universally in the data set has low self-information and is generally not very useful in the construction of hypothetical themes. The selected set of attributes obtained from this process is referred here as *the reduced attribute set* and is denoted as A_r.

Based on statistical evaluation, A_r is initially selected such that $f(a_i)\quad_1$ where $f(a_i)$ is the frequency of attribute a_i in the set of data. The most compelling reason for removing low frequency attributes is that it can be shown that such attributes will not be involved in the final theme set if certain criteria are required for selecting a theme.

Furthermore in practice, a set of known attributes can be disregarded in the construction process as they are descriptors that contain very little information specific to a domain. For example, it is desirable to remove insignificant words such as conjunctions (e.g. "and", "but", "or") and prepositions (e.g. "with", "but", "or") as text descriptors. These words are heuristically eliminated through the use of a database of such words or descriptors. Heuristic techniques may also be used to map several attributes that are known to be equivalent to a single new attribute. This is done with natural language when several words are known to be semantically indistinguishable or when words and their plurals are equivalent.

Attribute selection is performed using practical justification when the size of the attribute space is quite large and needs to be restricted to a reasonable limit. For example, in this implementation a limit of 1024 attributes is set and only the 1024 most frequent attributes remain in the description of the data set.
After the attribute selection process, the data are transformed using the reduced attribute set. It is possible that as a result, there exists no descriptor in describing some data. This is intuitively reasonable if for these data, no descriptor is used in constructing a theme.

In implementation, a binary vector can be used to represent the data in the set, involving only the attributes selected. To denote this process, the data $s = \{s_1\ s_2 \ldots s_m\}$ of m descriptors is transformed to a binary vector $X_i = (x_0, x_1, \ldots, x_{n-1})$ such that

$$x_j = \{ \begin{array}{ll} 1 \text{ if } s_k = a_j, & 0\ \ k\ \ m.\ 0\ \ j < n \\ 0 \text{ otherwise} & \end{array}$$

What this does is set x_j to 1 if a selected attribute a_j occurs in the data. Notice that the number of occurrences of an attribute within a datum is not relevant if more than once is observed.

In the remaining processes it will be valuable to make use of the frequency $f(x_i)$ of attribute x_i. The binary variable x_i's are sorted to consider the most frequent attributes first.

B. Theme Generation Process

The result of attribute selection and transforming the data using the reduced attribute set is a data set with attributes that may be involved in the construction of a theme. The next process is called *theme generation*. The aim of this process is to consider the complete set of all possible themes. This process considers the power set that is the set of all possible subsets. It is the Cartesian product of the reduced attribute set A_r with itself. If there are n variables (representing the attributes) in X_j then there are 2^n possible combinations of these variables. Normally, this set of possible themes is combinatorically explosive to be useful in practice. The following process is designed to generate a more manageable set of candidate themes.

C. Candidate Theme Generation and Theme Elinination

The set of candidate themes is constructed from the data set using the reduced attribute set which describes the data. It is a subset of the set of all possible themes. From this set, the final optimal theme set satisfying certain criteria will be identified to describe the whole data set. This process of constructing the set of candidate themes is similar to the algorithm of feature selection described in [Kittler 1986]. Here, we introduce a function called *matching index function.*

Definition 1. The *matching index* function M is a function

$$M : S \quad T \quad [0,1]$$

which maps a data in S and a theme in T to a real number in the interval [0,1], indicating the degree that the theme is observed in a data.

The index will return 1 when the data and the theme match perfectly, less than 1 for a partial match and 0 when there is no match. The function can be designed depending on the type of the data and how the data should be compared in the specified domain.

Definition 2. We define the *Cover Set* of theme t_i to be the subset of data in S such that $M(s,t_i)$ is greater than a certain threshold, $(s \quad S)$. The *Cover Set* of t_i will be denoted as $\quad (t_i)$.

The criterion function that will be used for evaluating a theme will be based on the number of data that a theme is observed in the data set, denoted as $\quad (t_i)$. By assuming that the themes generated as candidate themes must be observed at least once in the data set, four principles can be derived in constructing candidate themes:

Principle 1. The attributes with the highest frequency in the data will be considered first (statistical dependency).

Principle 2. Data are considered independently in selecting attributes in forming a candidate theme (structural dependency).

Principle 3. Candidate themes are constructed from as many suitable attributes as possible observed in the same data (weighted towards larger theme size*)*.

Principle 4. The union of the cover sets from the themes selected in a theme set should be sufficiently similar to the original data set (stable cost).

An algorithm based on these principles was developed, analogous to the selection algorithm using individual merit [Kittler 1986]. The merit of a theme for selection is evaluated as (t_i) and is used to order the themes in decreasing order of magnitude. This ordering is made possible by the fact that the selected attributes in A_t have previously been sorted by frequency of occurrences. Note that the maximum number of themes that can be constructed is limited by the number of attributes in describing a datum in the data set.

Algorithm 1: (Candidate theme generation based on individual datum). For each datum s in the set. Generate a theme t_i from the reduced attribute set of this datum. Let $D(s)$ is the number of themes that can be constructed. Compute the number of data in the data set that the theme t_i is observed, denoted as (t_i), i= 1, 2,..., $D(s)$. Rank the themes according to (t_i), in the order of decreasing magnitude. The selected themes are the first d(s) themes, such that $(t_i) > {}_2$, where ${}_2$ is a pre-defined threshold. Denote the generated candidate themes for this data as:

$$T(s) = \{ t_{i(s(1))} \mid i = 1, \ldots, d(s) \}$$

This procedure will be applied once for each datum s in the data set. A further pruning is done according to the data subset that a theme covers. If two themes cover the same data subset, then the theme with a smaller set of attributes will be disregarded. This is due to Principle 3 discussed previously.

D. Selecting the Optimal Theme Set for Describing the Data

The final process selects the optimal theme sets that can adequately describe the given data set. Considering that the given data set shares some common characteristics, the goal is to identify a set of themes such that the cover sets of the themes approximate the original data set. The result is that we can describe the data set by just referring to the attributes in theme set, without referring directly to the original data. This is the purpose of theme detection. However, the cover sets of the themes will not usually be exactly the same as the original data set due to random noise data. There are two criteria in evaluating a theme set in determining whether it can adequately describe the data set which the theme sets are generated from.

Definition 3. A theme set is a set of themes and is denoted as $\{ t_i \mid t_i = \{a_{i1}, a_{i2}, \ldots, a_{im}\}, i = 1, 2, \ldots, p\}$ where a_{ij} is a selected attribute that forms the theme and p is the size of the theme set.

Criterion 1. A theme set is of better "quality" to another theme set of the same size, if the set of the union of the cover sets of the themes in the set is larger.

Criterion 2. A theme set is of better "quality" to another theme set of the same size, if the cardinality of the union of the set of attributes which forms the themes is larger.

Selecting theme sets that have themes covering most of the original data, we can select the optimal theme set according to criterion 2 if all the theme sets considered are of the same size. That is, the preference is based on the combined number of attributes of the themes for a given theme set size. The selection process then selects theme sets for the various sizes. The theme sets of different sizes provide different interpretations or "views" of the data. Note that themes in the theme set of a large size may not cover a large number of data. Therefore, up to a certain limit, cases of large theme set size need not be considered.

Experimental Evaluation

A. Using a Small Set of Documents

This section demonstrates the method using a small set of documents from a web search example. When a web search is performed, the result is a set of documents, such that each document is described as text sequences representing titles, keywords or document descriptions etc. For example, consider the first ten 'hits' (web pages containing the keywords being sought) of a search for documents with keywords "data mining" and "research". After performing attribute selection, the data can be represented using the reduced attribute set as follows:

$s^{(0)}$={ data, mining, research, group }
$s^{(1)}$={ data, mining, research, home, projects, search }
$s^{(2)}$={ data, mining, research }
$s^{(3)}$={ data, mining, home, page }
$s^{(4)}$={ data, mining, research, projects, other, information }
$s^{(5)}$={ data, mining, research, analysis }
$s^{(6)}$={ data, mining, research, information }
$s^{(7)}$={ data, mining, page }
$s^{(8)}$={ data, mining, research, group, other }
$s^{(9)}$={ data, mining, search, analysis }

To illustrate the candidate theme generating process, consider $s^{(0)}$={data, mining, research, group}. The theme t_1 = {data} can be constructed. Since the attributes have been arranged in order of decreasing occurrence, the theme t_1 = {data} will be observed most frequently, which is indicated by $(t_1) = 10$, or in all the data. It is included in the set of candidate themes. Next, consider the theme of two attributes, t_2 = {data, mining}. Since $(t_2) = 10$, t_2 is also added to the set of candidate themes. Next, consider theme t_3 = {data, mining, research}. Since $(t_3) = 7$ which is still considered to be large and t_3 is added to the set of candidate themes. Next, consider theme t_4 = {data, mining, research, group} with $(t_4) = 2$ which is considered to be low and t_4 is rejected and the procedure halts. At this point, all the themes that cover the same samples can be removed. Since the two themes t_1= {data} and t_2={data, mining} cover exactly the same samples, the theme t_1={data} is removed, since it consists of fewer attributes. The algorithm will be repeated for the remaining samples. After considering all 10 documents, 11 candidate themes are found. After theme selection, one optimal theme set with only one theme in the set is found. It can be described by the theme t_2 = {data, mining, research, group} which describes sufficiently all the documents. For theme sets with two themes, two sets are found. Choosing the one with the largest number of unique attributes, the theme set can be described as: {t_2 = {data,mining,research,group}, t_1= {data,mining,research,projects,other}}. This theme set indicates that the whole set of documents can be intuitively described by two themes, one has the keyword "group" and the other has the unique keywords "projects" and "other".

B. Using a Large Set of Documents

The theme detection method as applied to web search data has been implemented in Visual C++ under WindowsNT™ 4.0 using a DDE link to Netscape Navigator 3.0. The AltaVista™ or Lycos™ search engine was used to generate the search results. The search engine provides a title and document description that was combined to provide a text sequence that represents each document. The user interface [Fig. 2] allows the user to enter the keywords of the search and also allows the user to select the search engine that should be used in the search.

After the user begins the search, the search string is formatted according to the search engine requirements in the form of a URL and submitted through a DDE link to Netscape Navigator. Up to 1000 search 'hits' are requested and the returned 'html' stream is parsed for document titles and descriptions. The text sequences are processed using the theme detection method and the results are displayed in real-time in a separate window. In this experiment, retrieving the search results from the engine takes much longer than the theme detection process itself. A trial search was done for the keywords "pattern recognition" and "research" which yielded the following theme sets. The optimal theme sets of different sizes covering 80% of the retrieved documents are found: (1) Theme set of 1 theme: {pattern, recognition} with two unique attributes; (2) Theme set of 2 themes: {{pattern,recognition}{pattern,recognition,research,processing}}, with four unique attributes; (3) Theme set of 3 themes: {{pattern,recognition} {pattern,recognition,research,image} {pattern, recognition, research, processing} } with five unique attributes; (4) Theme set of 4 themes:

BEST COPY AVAILABLE

{{pattern,recognition} {pattern,recognition,computer,neural} {pattern,recognition,research,information} {pattern, recognition, research, image, processing}} with eight unique attributes. From the resulting theme sets, theme set 4 has the largest number of unique attributes and contains the most descriptive information about the retrieved documents. They roughly divided the search resulting documents into four groups. Notice that theme set 1 and 2 simply include the search keywords and are the intuitive obvious themes of the search results.

References

[Kittler 1986] Kittler, J. (1986). Feature Selection and Extraction, in *Handbook of Pattern Recognition and Image Processing*, Ch.3, (eds.) T.Y. Young and K.-S Fu, Academic Press, 59-83.

[Jain 1988] Jain, A.K. and Dubes, R.C. (1988). *Algorithms for Clustering Data*. Advanced Reference Series, Prentice Hall.

[Chiu 1986] Chiu, D.K.Y. and Wong, A.K.C. (1986) Synthesizing knowledge: A cluster analysis approach using event covering. *IEEE Trans. on Systems, Man and Cybernetics*, 16 (2), 251-259.

[Lehnert 1994] Lehnert, W.G. (1994), Information extraction as a basis for high-precision text classification. *ACM Transactions on Information Systems*, 296-333.

[Zaiane 1995] Zaiane, O.R. and Han, J. (1995), Resource and Knowledge Discovery in Global Information Systems: A Preliminary Design and Experiment, *Proc. KDD-95*, 331-336.

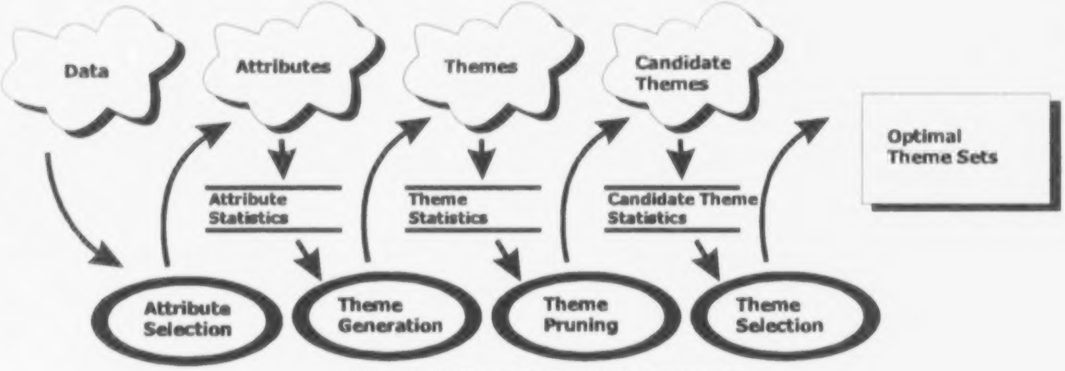
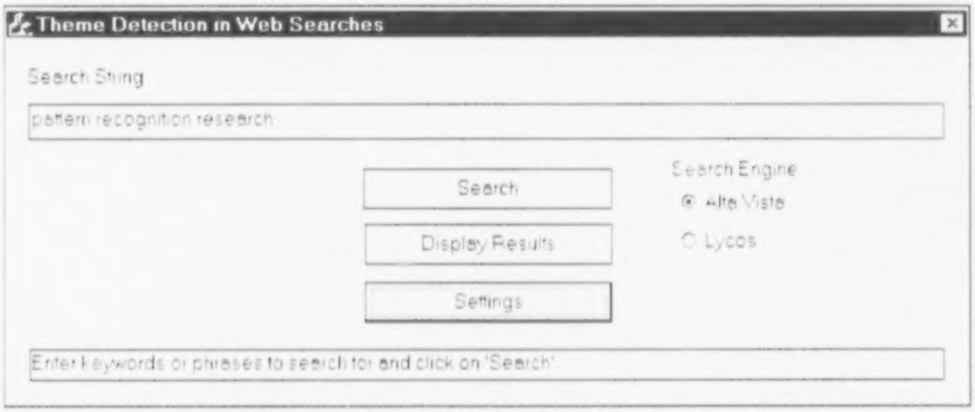

Data
Attributes
Themes
Candidate Themes
Optimal Theme Sets
Attribute Statistics
Theme Statistics
Candidate Theme Statistics
Attribute Selection
Theme Generation
Theme Pruning
Theme Selection

Figure 1: The theme detection process


Theme Detection in Web Searches
Search
Display Results
Settings

Search String

pattern recognition research

Search Engine

● Alta Vista

○ Lycos

Enter keywords or phrases to search for and click on 'Search'

Figure 2: The user interface for web search theme detection

BEST COPY AVAILABLE

Building a Configuration Assistant on the Web

Antonio Cisternino
Giuseppe Attardi
Maria Simi

Dipartimento di Informatica, Università di Pisa, Italy
{attardi, cisterni, simi}@di.unipi.it

Abstract: We describe a general tool for developing configuration applications running on the Web. Starting from a declarative description of the basic items to be chosen for the configuration and of the configuration constraints, the tool generates the HTML files for user guidance and the Java code for constraints checking. An interactive assistant for compiling and submitting plans of study has been built with the tool and deployed at our university.

1. Introduction

In the terminology of expert systems configuration systems are a subclass of design systems, whose task is to *assemble* a set of predefined objects that satisfy a given set of problem specific constraints [Hayes-Roth, Waterman & Lenat 1983]. Examples of configuration problems are computer equipment configuration (XCON [Barker & O' Connor 1989]), software configuration, timetables generation and scheduling.

A configuration task is generally rather complex since it involves coping with many and interacting design decisions, whose consequences cannot readily be assessed, and constraints of different nature. Configuration problems are therefore a challenging domain for expert system technologies.

For simpler configuration tasks we can envision interactive *configuration assistants* which operate by guiding the user step by step through the available design decisions by exploiting their knowledge of the domain constraints and of the constraints deriving from previous choices. Configuration tasks that are amenable to this simplified vision are for example plan of study compilation or the assembling of any coherent system built from a catalogue of components, like a complex piece of furniture (a kitchen furniture for instance) or a personal computer.

Configuration assistants of this kind running on the Web have a lot of potential in many fields, including electronic commerce, for the wide availability and the possibility of remote access and use. The system does not need to be installed on the users computer in order to be used and portability problems need not to be addressed. This is also an application domain where a Java solution [Gosling 1996] has clear advantages over a server based solution where the client interacts via a CGI interface: all the job of the configuration assistant can be done locally at the client's side by downloading the necessary Java code; communication with the server can be reduced to tasks such as user validation, statistics gathering or archiving.

In this paper we present the main ideas behind a general model for configuration and describe a tool for developing specific configuration assistants running on the Web. The configuration model can be characterized as *process based*, in contrast to *product based* configuration models, since the aim is guiding the user step by step trough the configuration process rather than starting from a high level description of the product to be configured. A configuration application is generated starting from a high level description of the basic components and the constraints expressed in a declarative form. The HTML files for user guidance and the Java code for constraints checking are automatically generated from this high level description.

In order to demonstrate the use of the general tool we will describe the generation and use of a Web assistant for plan of study compilation and submission (CompAss) which has been developed for the Faculty of Letters and Philosophy of the University of Pisa.

We will conclude by discussing limitations of the current system and plans for future developments.

2. Building a Configuration Application

A configuration product it built from a set of predefined *basic items* that the user can select, whose combination has to satisfy a set of domain specific configuration constraints. A *configuration domain* is defined by the complex of items and constraints specific to a configuration application.

The *process oriented* configuration model we use relies on a directed acyclic graph called *choice graph*; each node of the graph corresponds to an available user alternative in the configuration process and defines a set of corresponding constraints, typically the items required as a consequence of the choice; successor nodes in the graph correspond to subsequent choices in the configuration process. A *configuration* is a subset of the basic items, and a set of intermediate user choices, that match a given set of constraints. A configuration is *valid* only if the user choices define a path from the root to a leaf, and the selected items match all the constraints associated to the nodes in the path.

To build a configuration application we specify, in a declarative form, the various aspects of the configuration domain; in particular the choice graph, with associated configuration constraints, is defined in a declarative constraint language and made available in a *constraint file*.

2.1 The Constraint Language

The constraint file, defining the choice graph, is the heart of a configuration application; in fact the *validator* module of the application interprets the constraints defined in this file to test the configuration. The constraints are defined in a special declarative language designed for this purpose.

The language reflects the structure of the choice graph, which is a natural way to think of the configuration process in many application domains. The language supports the definition of two types of blocks: *list blocks* and *choice blocks*. A list block is simply a way to define a group of items so that it can be referred by name. A choice block corresponds to a node in the choice graph and defines the various constraints associated to the node. A choice block has the following structure:

```
Name of the block {
 [item_needed_1,
  item_needed_2,
  ...],
 [item_pres_1, item_pres_2, ..., item_pres_n]=>            /* 1 */
    [item_needed_n+1, ..., item_needed_n+h]
 [item_choice_1, ..., item_choice_k, #ref](7),             /* 2 */
 [#ref_1, #ref_2](2+),                                     /* 3 */
 [#ref_3, item_choice_k+1](1-)                             /* 4 */
 CHOICE(Block_1, ..., Block_m)                             /* 5 */
} f1(32), f2(31);                                          /* 6 */
```

In a block description, we can find items that are necessarily needed for the block and items that are needed depending on the configuration state, i.e. the presence of other items. For example line 1 says that the items on the right of the `'=>'` operator are to be included in the block only if the items on the left are present in the current configuration. The condition for inclusion can also be a combination of logical operators (AND, OR, NOT). Line 2 is an example of a construct that prescribes the selection of a number of items out of a list of items; in particular the example says that exactly seven items of the configuration must be selected from the given list. Lines 3 and 4 are similar with different number restrictions: in the first case *two or more* items are required, in the second case *at most one* item is accepted. Line 5 defines the successors of the current node in the choice graph, i.e. the available choices at this level. The operator '#' is the way to include all the items belonging to a defined *list block*; for example `#ref` refers to a block named `ref`, which could be defined as

`ref[item_1, #ref_2, ..., item_n]`

and in turn could contain references to other list blocks.

The choice block defines the items required for the node in the configuration. The type of standard controls that are generated concern the admissibility of items (i.e. answers the question "is it correct that this item is in the current configuration"?) or the presence of items (i.e. answers the question "is this required item

BEST COPY AVAILABLE

present in the current configuration"?). These constraints result in a set of built_in control functions such as *Nec* (for necessary items), *Atleast, Atmost, Exactly* (for numerically restricted selection from a list of items).

Other kinds of constraints which are often needed, such as "The cost of the configuration must be at most XXX$", are implemented by *custom constraints functions*, which are typically application dependent: these can be defined by the user or supplied by a library. Custom constraints functions appear at the end of a choice block (like in line number 6 above) and are usually aggregate boolean functions which apply to all the items in the block and successor blocks.

2.2 Item Structure, Item Database, and Custom Constraint Functions

In addition to the constraint file, three additional data files, in human readable form, are necessary to build a configuration application: the *item structure*, the *item data*, and the *custom constraints functions*.

The item structure file contains an item definition (similar to a `struct` of the language C). The items themselves are described in a text file according to the defined item structure. At least two fields are required in an item description: the *code field* and *the name field*. The code field is a fixed size field which plays the role of an access key for the item and is fundamental for passing parameters. The name field is a variable size description of the item to be used by the applet at runtime in communicating with the user. Each item also contains a description in HTML that is displayed in the documentation frame on user's request and possibly other application specific fields.

For the definition and enforcement of domain specific constraints the user can define special custom functions, in addition to the standard constraints resulting from the constraint file described above; these functions are boolean tests on the current configuration state and are defined in the custom constraints file.

2.3 Generation of a Configuration Application

A configuration application is generated by using a compiler which takes as input the data files described above. The compiler is divided in two modules: C1 and C2 [Fig. 1].

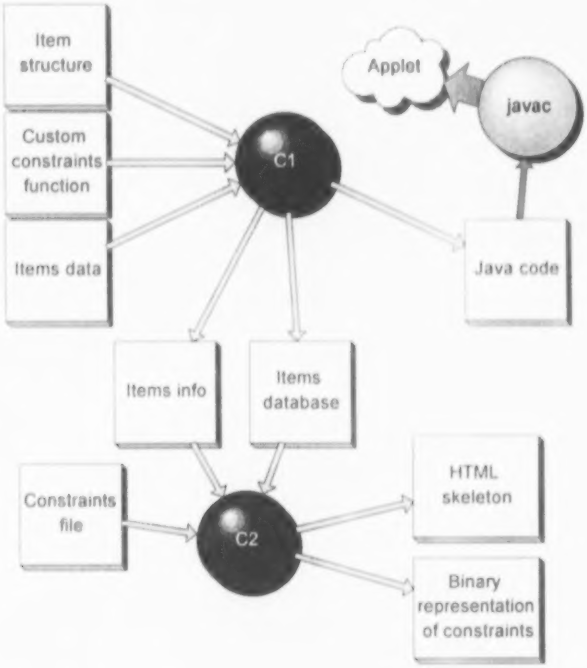


Figure 1: Generation of a Configuration Application

BEST COPY AVAILABLE

The first module is needed for generating the Java code for the applet and the items database. The second module of the compiler generates a binary representation of the constraints and the HTML files for documentation and user guidance.

More specifically, the module C1 of the compiler takes as input the three data files described above (the *item structure*, the *items data*, and the *custom constraints functions*) and produces three files which are used in the second step of the compilation process: a Java program, information about the items in HTML form and the binary version of the items database. The generated Java applet depends on the configuration domain only for the item structure and the custom constraints functions; these elements are Java classes generated by the compiler and later combined with the rest of the applet. The applet will also use the items database and a binary representation of the configuration constraints. The Java code includes the *item Java class* and the *custom constraints functions Java classes*. The item class is the class that describes the format of the items database and offers to the configuration applet a set of methods for reading, writing and accessing item components. The custom constraints functions classes are a Java version of the custom constraints functions. These classes are managed by a Java class that maps the function calls to the proper functions.

The module C2 of the compiler takes as input the constraints file, the items information and the items database. It generates HTML files and the binary version of the constraints.

The HTML files generated are to be used in user interface of the configuration assistant. In particular a set of HTML skeleton files are generated out of the choice graph: for each node in the graph a file is generated with selection icons for the items in the node and hypertextual links to the items descriptions. The file also contains a few lines of text which synthetically describe the node constraints (for example "Choose at least three items out of the following:"), which can be enriched with additional text deemed useful to guide the user during the configuration process. In addition the file contains choice icons and hyperlinks to other HTML files in correspondence of available choices.

A compact binary representation of the choice graph is also generated by C2 and it is the primary data structure used by the applet for checking the validity of a configuration.

3. Communication with the Server

One of the major issues in the use of Java applets for building applications on the Web is security. The *Security Manager*, i.e. the Java class that defines security policies for Java, prevents the applets from doing I/O on the client local disc and allows opening sockets only with the Web host from which the applet has been loaded. These limitations make very difficult to write applications that use persistent data.

Our solution is to use a special server on the Web host that listens to a given TCP port: the applets open sockets to this server on the specified port and use the server for saving data. Java has convenient facilities for communicating via sockets and Object serialisation is useful for sending Java objects across the Web.

The server is written in Java and allows different kinds of clients: configuration clients but also server console clients. The configuration clients are the applets running in a configuration application. The server console clients are Java standalone applications that allow remote monitoring of the server. Through the server console the user can monitor in real-time configuration clients, displaying the Internet hosts with open configuration connections, and save statistics on the use of the system.

The server also provides local printing capabilities generating and sending back HTML pages with the data provided by the client in the required format; the user can print the content of the generated page with the regular print button of its browser. The server can also store these data in a database for later use.

4. CompAss: a Configuration Assistant for Plans of Study Compilation

CompAss (COMPilazione ASSistita di piani di studio) is a system which assists students in the task of producing a plan of study. CompAss and its associated support tools have been developed in the context of a pilot project for the Faculty of Letters and Philosophy of the University of Pisa.

Plans of study approval is a time consuming job for all the courses of study in the faculty, due to the high number of submissions each year (around 3000) and the high rate of incorrect submissions. One of the requirements was that students could use any computer located in the various departments of the faculty to compile plans of study; data had to be collected in one single place for archival. The Java solution was the obvious choice and offers additional advantages such as the possibility of using the system from home.

The Web page of the CompAss configuration assistant is vertically divided in two parts [Fig. 2]. The right part contains the *navigation frame* with related title bar and navigation buttons, the *help frame*, and the *documentation frame*. The left part contains the *configuration frame* and an *application specific tool bar*.

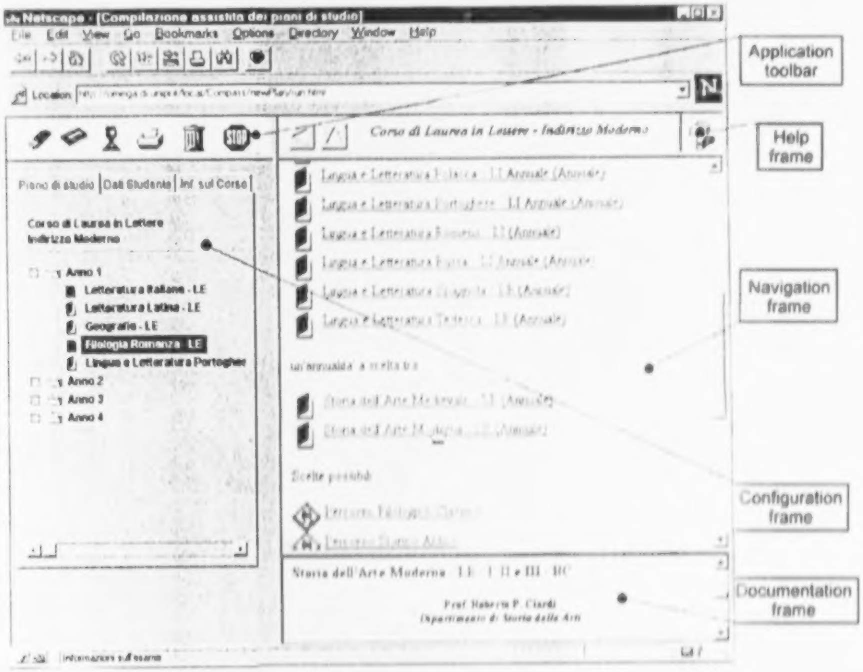

Netscape - [Compilazione assistita dei piani di studio]
File Edit View Go Bookmarks Options Directory Window Help
Location:
Corso di Laurea in Lettere - Indirizzo Moderno
Piano di studio
Dati Studente
Inf. sul Corso
Corso di Laurea in Lettere
Indirizzo Moderno
Anno 1
Letteratura Italiana - LE
Letteratura Latina - LE
Geografia - LE
Filologia Romanza - LE
Lingua e Letteratura Portoghese
Anno 2
Anno 3
Anno 4
Scelte possibili
Storia dell'Arte Moderna
Informazioni sull'esame
Application toolbar
Help frame
Navigation frame
Configuration frame
Documentation frame

Figure 2: The User interface of CompAss

The navigation frame is a HTML frame displaying a normal hypertextual document; it displays the available choices together with any informative text deemed useful to guide the user to do the right choices during the configuration process. Hyperlinks are associated to configurations items and, when clicked, make a description associated to the item appear in the documentation frame.

Special icons associated to choice points and to items are used to perform configuration actions: intermediate choices or item selections; when these icons are selected they send messages to the configuration program, the *validator*, which is a Java Applet associated to the configuration frame.

The configuration frame on the left contains the Java applet which manages the configuration. The applet receives input by direct interaction in its client area (handled through events in the AWT) or by selection of special icons in other frames (the navigation frame and the tool bar frame). Whenever a configuration action is performed the applet reacts by checking the current partial configuration, accepting the change or prompting the user if any configuration constraints is violated.

Tool icons in the toolbar denote general utility or application specific actions available to operate on the partial configuration displayed in the configuration frame (i.e. item deletion, final configuration validation, abortion of the configuration process, printing or submission of the final configuration).

One difficult technical problem was to allow interaction between HTML pages and the Java applet which collects and stores the plan of study. An HTML page was the appropriate way to describe the courses of study and provide documentation and guidance to the student in filling the plan. It would have been nice to allow the user to pick up a course (through its title or an icon representing it) and drop it in the plan. Drag-and-drop operations between HTML and applets are not currently supported. Therefore we had to resort to a solution where selection of an item is performed by clicking on the corresponding icon. However filling a page with dozens of Java applets (one for each course) would have been unfeasible, bogging the browser. The solution has been to use JavaScript to post the events of icon selection to the configuration applet. With this solution the

BEST COPY AVAILABLE

interface can exploit all the power of the HTML language and standard browsing capabilities, while still allowing user interaction with the Java programme.

In this configuration application the basic items are all the courses offered by the faculty; the constraint file implements the rules for plan of study formation; it includes a choice graph where nodes correspond to choices such as the course of study, the orientation, the field of specialisation and so on, together with the necessary constraints. A configuration is a legitimate plan of study, i.e. a list of courses which a student plans to take, fulfilling all the requirements imposed by the faculty.

The official submission of the plan must be done on paper because it requires a signature by the student. Our current solution is that the plan is printed locally, after completion and verification by CompAss, and automatically sent to the server and registered in a temporary area. When the student submits the plan to the secretary office, the plan is retrieved and transferred to the archives of submitted plans. CompAss saves a lot of work for secretaries who previously had to type in the plans from the paper forms submitted by students and eliminates the routine work of the faculty committees which had to verify and approve the plans.

The plan of study manager running on the server accepts communications from several CompAss clients, receives data from plans of study, generates HTML pages, stores data in a database, and gathers statistics on the number of users and on the pattern of use of the system.

CompAss can be seen at the Web address "http://omega.di.unipi.it/local/Compass/start.html".

5. Conclusions and future work

We have described a general tool for generating configuration assistants; the strategy works well in the specific configuration domain of plan of study compilation, but we believe that other configuration applications are amenable to this simple paradigm. More experimentation is however needed to exactly define the range of applications and to come out with a general enough configuration language.

We plan to enhance the configuration language by including a language-level specification for defining the structure and topology of the product to be configured, which largely depends on the application domain. This will also influence the display of items in the configuration frame.

We foresee also some improvement due to advances in the Java technology: with the new version of HotJava, provided by Sun Microsystems, we will be able to exploit a new capability for communication between HTML and the Java applet. For the same purpose, we also plan to write a new interface that uses the Netscape live-connect system.

6. References

[Hayes-Roth, Waterman & Lenat 1983] Hayes-Roth, F., Waterman, D.A. & Lenat, D.B. Eds. (1983). *Building Expert Systems*, Addison Wesley Publishing Company, Reading, Ma.

[Barker & O' Connor 1989] Barker, V.E. & O'Connor, D.E. (1989). Expert Systems for Configuration at Digital: XCON and beyond. *Communications of the ACM*, 32(3):298-318.

[Gosling 1996]Gosling, A.J. (1996). *The Java Programming Language*. Addison Wesley Publishing Company, Reading, Ma.

Acknowledgements

We are grateful to the Faculty of Letters and Philosophy of the University of Pisa, and prof. Mirko Tavoni in particular, for sponsoring the project, to the CISIAU (the computing centre of the Faculty) and its technical director dr. Paolo Pisanti for providing the necessary technical support and to the people who helped us to understand the rules for plans of study (prof. Letta and prof. Scalfati). The student Alessandro Pelosi also contributed in early stages of the project.

WebSteps, a Tool for Synthesizing and Managing Web-based Business Processes

Jürgen Coppens, Oscar Figueiredo, Roger D. Hersch, Benoit Gennart,
Ecole Polytechnique Fédérale, Lausanne, Switzerland
{Oscar.Figueiredo|Roger.Hersch|Benoit.Gennart}@di.epfl.ch

André-R. Probst, Maia Wentland Forte
HEC-INFORGE, University of Lausanne, Switzerland

Frédéric Vianin, Exel&Scopus, Lausanne, Switzerland

Abstract: This contribution introduces WebSteps, a tool for easily creating and running a commercial Web server. WebSteps executes a business process described by a state chart. The description of the state chart, also called business scenario, is analyzed by a program running on the server (WebSteps engine). The WebSteps instructions associated with the current state are executed. The WebSteps language describes actions to be carried out within a state by only four types of executable instructions (operations, conditional operations, HTML page posting, change state instructions). We apply WebSteps to create a Web server for apartment renting. WebSteps is a flexible tool for synthesizing commercial Web servers since it combines for the purpose of a business process HTML information pages, customized HTML forms for the interaction with clients and accesses to the server's databases.

1.1 Introduction

In today's global market, all corporations, big and small, need to constantly be on the outlook for new clients. The World Wide Web offers an excellent platform for corporations to present and sell their goods and services. While large corporations have the means to set up sophisticated Web servers, the situation is very different for small businesses. They often don't have the resources (financial, time, know-how etc.) to develop the Web-based technologies they would need to automate their business processes.

Constructing efficient server-based Web applications requires to master techniques beyond non-programmers capabilities, such as the Hypertext Markup Language (HTML), the Common Gateway Interface (CGI) and a programming language, such as C++, Perl or Java. Business processes, while often showing some similarities, vary from one business to another. Most commercial databases offer Web extensions for formulating queries, and receiving results within HTML pages. For the purpose of information display, different applications such as Microsoft FrontPage [Microsoft 1997], Adobe PageMill [Adobe 1997] or MapEdit [Boutell 1997] greatly simplify the creation of HTML pages. These tools assist the creation of individual pages but do not provide help for setting up a Web server for business processes. To simplify the development of a commercial Web server, a tool that manages the interactions between databases, HTML information pages and HTML interactive forms must be provided.

Previous research has focused on schema based approaches for HTML authoring. Interesting efforts include the Relationship Management Case Tool [Diaz et al. 1995] which is based on a data model describing the architecture of the Web site. The Hypertext Structure Description Language [Kesseler 1995] gives special attention to schema evolution thus facilitating non-trivial updates. These approaches assist the creation of presentation oriented servers. In contrast to these tools, WebSteps offers a dynamic management of the interactions between the databases, HTML information pages and HTML interactive forms. As the client moves along the subsequent states of a business process, WebSteps interacts with the databases and customizes HTML pages as needed at one particular step of the business process. The WebSteps state-based approach is particularly well suited for commercial Web servers as the different steps of a business process need the execution of specific operations.

WebSteps is based on the description of the business scenario that the site owner finds appropriate for his business. The script of the business scenario is analyzed and the necessary operations are carried out. The proposed tool is easy to use and has been applied for creating a prototype site enabling the reservation and renting of apartments. Our aim has been to create a tool as simple as possible, enabling a web site designer to describe his business process with extremely simple instructions comprising only a few basic types.

1.2 Modeling a Business Process by a State Chart

Each business process can easily be broken down into a number of distinct steps. For example: a database query to determinate available products meeting specific criteria, the presentation of the products, the reservation, the payment etc. Although these steps vary according to the type of the business, business processes can be characterized by such step sequences. A state chart is an intuitive way to represent the different steps of a business process and the interaction between those steps [Glintz 1995].

The function of a state in a state chart depends on its definition [Rumbaugh et al. 1991]. A state can be an operation (for example a database query) or a group of operations or even an instant between operations or groups of operations. In WebSteps we consider a state to be made of a sequence of instructions, only one of those instructions being an HTML page posting action (*makepage*). The *makepage* enables the client to interact with the server, thus influencing future developments of the business process. At the end of any given state, state changing instructions (if *condition* goto *new state*) will be found, permitting to access the next state of the business process. Associating only one HTML page per state makes the description of the business process more intuitive since a user with no programming experience will tend to associate a step in the business process with the visualization of new information by the client. A partial view of a state chart is depicted in [Fig. 1] while [Fig. 2] shows the associated HTML pages.

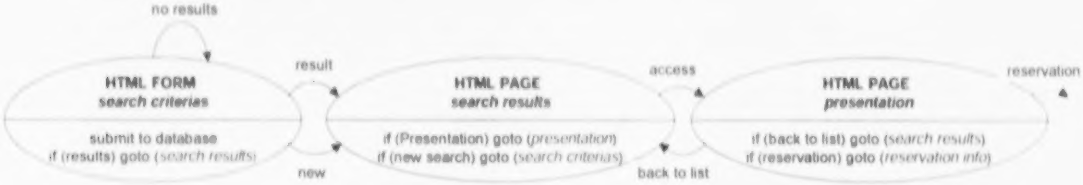

no results
result
access
reservation
HTML FORM
search criteries
HTML PAGE
search results
HTML PAGE
presentation
submit to database
if (results) goto (search results)
if (Presentation) goto (presentation)
if (new search) goto (search criterias)
if (back to list) goto (search results)
if (reservation) goto (reservation info)
new
back to list

Figure 1: *Example of a state chart describing the search criteria and presenting the results*

1.3 WebSteps Design

WebSteps has been developed in a PC environment, under Windows NT. The engine and the different subroutines have been written in Perl [Perl 1997] and are therefore portable across platforms. WebSteps itself is made of six different elements: the scenario description file, the engine, the Perl subroutines, the databases, the HTML files for display and the HTML forms [Fig. 3].

1.3.1 The Scenario Description File

A person desiring to synthesize a commercial Web server creates a scenario description file written with WebSteps instructions describing the state chart of the application. These instructions are interpreted by the WebSteps engine. WebSteps works in the same way as a state machine. In each state a number of instructions are executed: operations corresponding to executable Perl procedures, a single page posting instruction (*makepage*) which enables the interaction with the end-user, conditional instruction executions and a state change instruction.

BEST COPY AVAILABLE

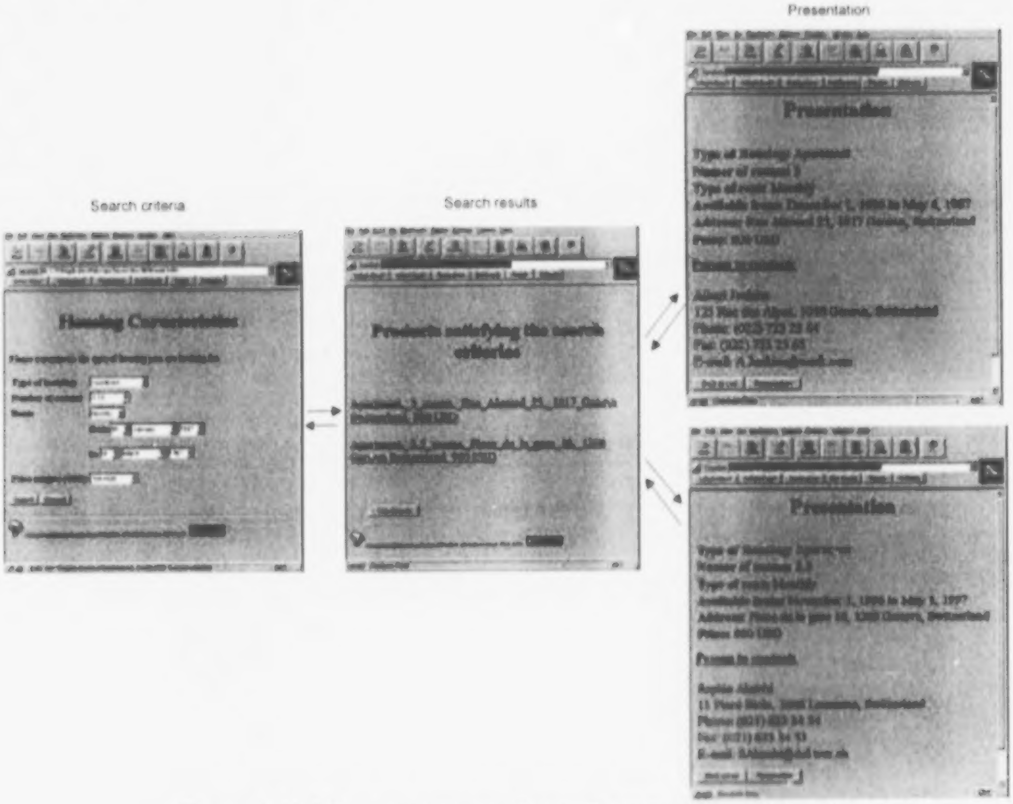

Search criteria
Housing Characteristics
Search results
Products satisfying the search criteria
Presentation
Presentation
Presentation

Figure 2: HTML pages corresponding to the state-chart of [Fig. 1]

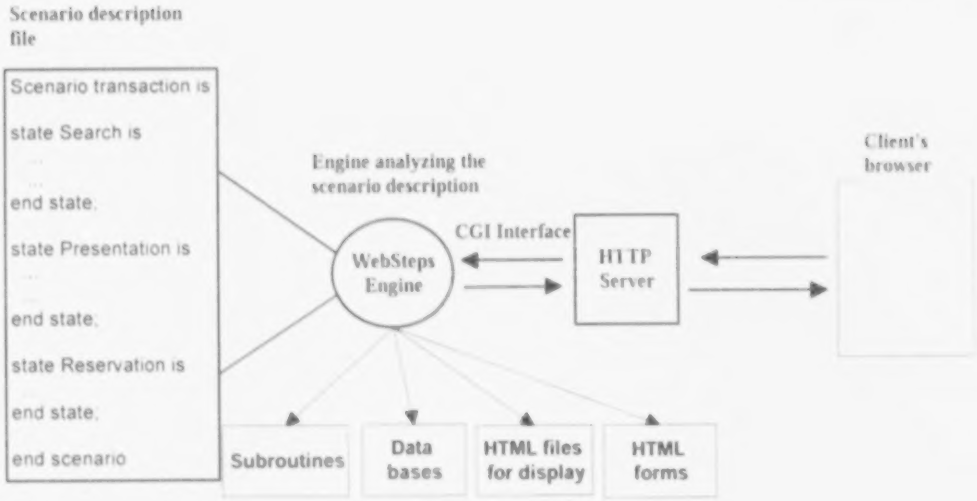

Scenario description file
Scenario transaction is
state Search is
...
...
end state;
state Presentation is
...
...
end state;
state Reservation is
end state;
end scenario
Engine analyzing the scenario description
WebSteps Engine
CGI Interface
HTTP Server
Client's browser
Subroutines
Data bases
HTML files for display
HTML forms

Figure 3: Components of WebSteps

BEST COPY AVAILABLE

Only four types of instructions are needed:

- Synthesis and posting of an HTML page (*makepage*)
- The execution of a single instruction (*operation* opName)
- *If* (condition*) do (*operation/makepage*)*
- *If (*condition*) goto (*state label*)*

Since a state is made of a group of instructions, the above primitive language is sufficient for describing the different states of the state machine and the transitions between these states. Since HTTP is a stateless protocol [Fielding et al. 1997][Berners-Lee et al. 1996], no HTTP transaction is defined in terms of the transactions that precede it. Therefore, a client ID and a state ID must be passed through the HTTP protocol at each step of the business process. After posting a page to a client, the WebSteps engine can be restarted either when a query is submitted by a client or when the client clicks on a link. It is therefore necessary that all HTML pages and links within those pages contain the identifier of the state to which they are attached.

1.3.2 The WebSteps Engine

The engine is the central part of the application. It is responsible for three different operations: identify the instructions that need to be executed, parse those instructions and finally execute them.

Identifying the instructions that need to be executed

Within a state, instructions can be separated into two groups, depending upon whether they take place before or after the interaction with the client (*makepage*). The instructions up to the *makepage* are executed first, then the connection between the server and the client is suspended. It is re-established after receiving the answers from the client. The instructions after the *makepage* are then executed [Fig. 4].

Parsing instructions

Following a client response, the instructions to be executed are first identified and then parsed [Fig. 5]. The following information must be obtained:

- The type of the instruction:
  - *operation* which leads to the execution of a subroutine written in Perl.
  - *makepage* which will create the page, send it to the client and enable him to interact
  - *if* (condition) *do* (*operation/makepage*): conditional operation or *makepage* execution
  - change state instruction (*if .. goto ..*) permitting to jump to the next state of the business process
- The name of the instruction itself, following the instruction type, for example *GetTime* [Fig. 4].
- Input variables: these variables have values which will be used for the execution of that instruction.
- Output variables: these will contain the results of the execution of the instruction.
- Once this information has been analyzed, the instruction can be executed.

Executing the instruction

WebSteps instructions are translated into valid Perl source code. This gives WebSteps great flexibility since additional instructions can be easily implemented by writing the associated Perl routines without altering the underlying structure. A new instruction, when used in the scenario, is automatically recognized and executed by the engine.

1.3.3 The Subroutines

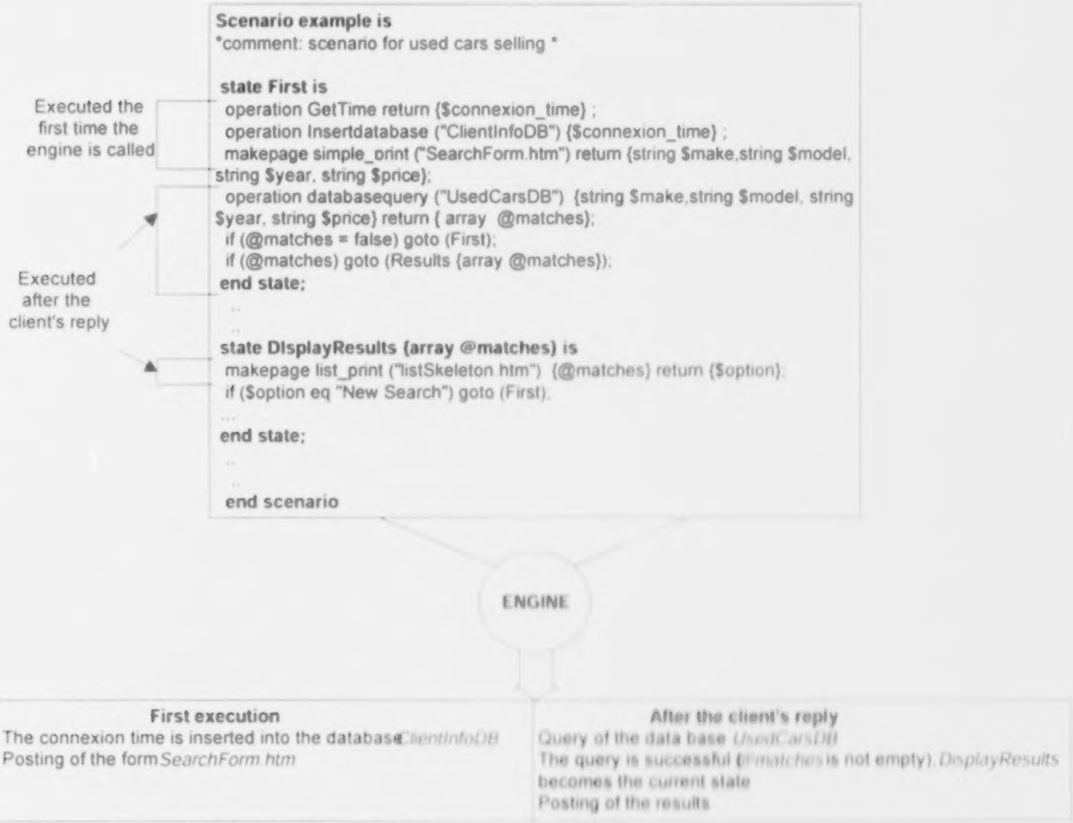
Each subroutine relates to one type of instructions. The following instruction types are currently supported:

- The *makepage* instruction enables various HTML page postings. It can be anything from the simple posting of an existing HTML page to inserting personalized information into an HTML skeleton. These skeletons, first developed with a conventional HTML editor are then added markers for the insertion of information. This enables the values of the input variables passed to *makepage* to be inserted into an existing HTML frame.

operations specify executable procedures written in Perl. These Perl procedures are selected accordingly to the operation name. Currently implemented standard operations include database queries, sending e-mails and saving information both in the databases and in the server's file system.

State change instructions enable the business process to move forward by moving into the next state when all operations relative to the current state have been executed satisfactorily.

Conditional execution of (*operation/makepage*) enable their execution under precise conditions.



```
Scenario example is
"comment: scenario for used cars selling "

 state First is
  operation GetTime return {$connexion_time} ;
  operation Insertdatabase ("ClientInfoDB") {$connexion_time} ;
  makepage simple_print ("SearchForm.htm") return {string $make,string $model,
string $year, string $price};
  operation databasequery ("UsedCarsDB")  {string $make,string $model, string
$year, string $price} return { array  @matches};
  if (@matches = false) goto (First);
  if (@matches) goto (Results {array @matches});
 end state;
  ..
  ..
 state DisplayResults {array @matches} is
  makepage list_print ("listSkeleton.htm")  {@matches} return {$option};
  if ($option eq "New Search") goto (First);
 ...
 end state;
  ..
  ..
  end scenario
```

Executed the first time the engine is called
Executed after the client's reply
ENGINE

First execution	After the client's reply
The connexion time is inserted into the database *ClientInfoDB* Posting of the form *SearchForm.htm*	Query of the data base *UsedCarsDB* The query is successful (*@matches* is not empty). *DisplayResults* becomes the current state Posting of the results

***Figure 4**: Example of a WebSteps scenario*

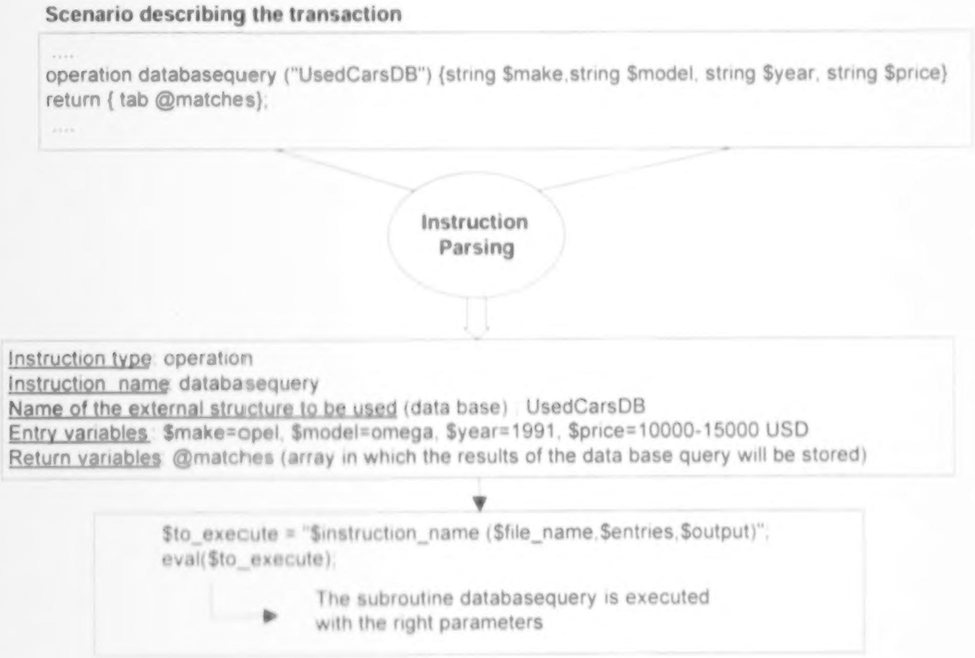
1.3.4 The HTML Files

Websteps considers two different types of HTML pages: completely designed pages that will be posted without any changes, and HTML skeletons into which personalized information will be inserted for each client. The second type of pages will typically be used to present the results of a database query, or the information concerning a previous reservation.

HTTP being a stateless protocol, a client ID and a state ID must be passed to the engine after each interaction with the client. The state ID is necessary so the engine can identify the current state in the scenario and thus the next instructions needing to be executed. It is inserted into HTML forms using hidden fields. The client ID is necessary because HTML pages are customized. HTML pages must thus be linked to the client to avoid conflicting situations with multiple clients. This has been achieved using HTTP Cookies [Kristol & Montuli 1997].

1.4 Example Application: Apartment Renting

A prototype site enabling the reservation and renting of apartments has been created using WebSteps. The state chart of the business process is represented in [Fig. 6].



```
....
operation databasequery ("UsedCarsDB") {string $make,string $model, string $year, string $price}
return { tab @matches};
....
```

Instruction type: operation
Instruction name: databasequery
Name of the external structure to be used (data base) : UsedCarsDB
Entry variables: $make=opel, $model=omega, $year=1991, $price=10000-15000 USD
Return variables: @matches (array in which the results of the data base query will be stored)

```
$to_execute = "$instruction_name ($file_name,$entries,$output)";
eval($to_execute);
```

The subroutine databasequery is executed with the right parameters

Figure 5: *Parsing one instruction of the scenario description file*

In this example the possibility has been given to pre-reserve an apartment before the final reservation. Once a client makes a pre-reservation he is given a client ID This enables him to transform this pre-reservation into a final one: when a client accesses the site (*welcome* state), he can confirm a previous reservation (ID number of reservation fetched from the cookie or given interactively) or begin a new search. After confirmation, the reservation in the database is confirmed. The proposed reservation scheme [Fig. 6] is only one of many possible solutions. It can be easily extended by adding further states and transitions in the business scenario.

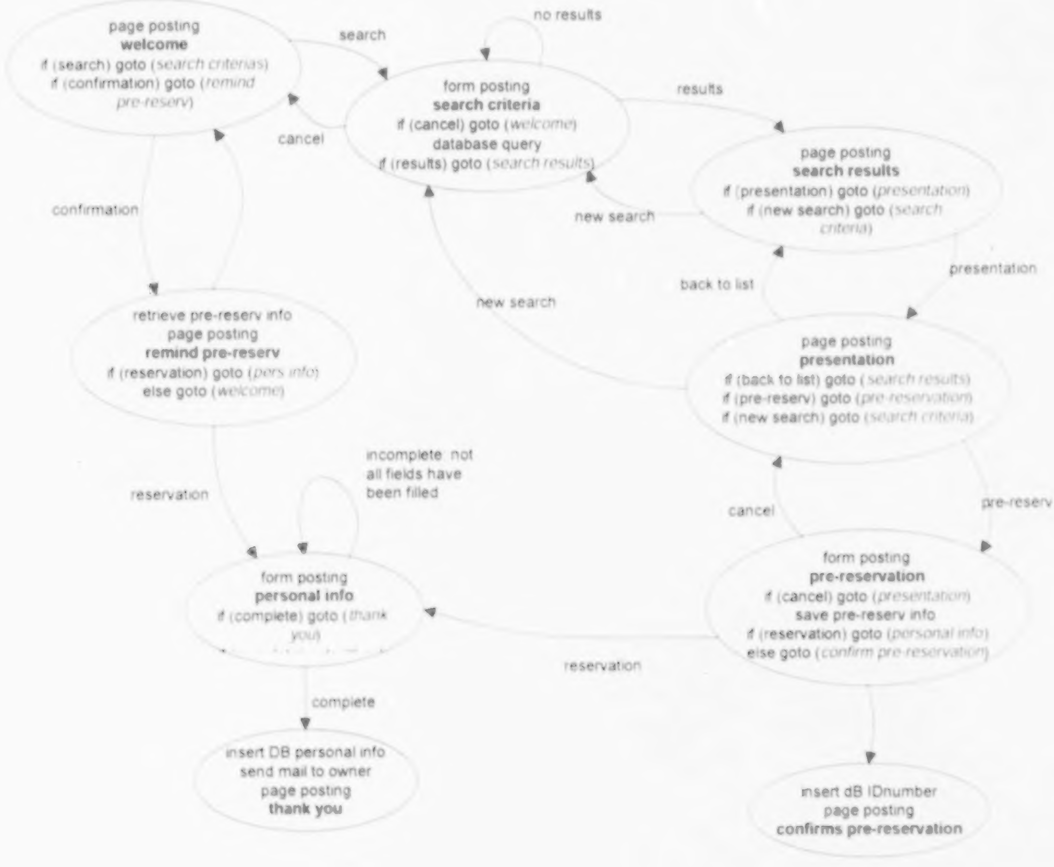
1.5 Conclusion

The WebSteps state chart based approach for synthesizing commercial Web servers offers several advantages. The description of a business process by a state chart is intuitive, since it is representative of the different steps leading to the completion of the business process. Furthermore, by associating a unique HTML page posting action with each state we have a well-defined paradigm for segmenting a business process into states. WebStep itself is based on a simple concept, enabling to easily design Web servers functioning like a state machine. This opens avenues to interesting operations such as overlapped database queries.

WebStep's biggest advantage is the customization facilities it offers. The designed site can be made more sophisticated by adding new states in the sate chart, making use of existing instructions. If necessary, new instructions can be created by writing the corresponding Perl routines. The tool that has been created is extremely simple and its associated language is easy to manipulate since four types of instructions are sufficient to describe the state chart associated to a business process.

Further developments aim at enabling non-programmers to create a business scenario by making use of an interactive graphical user interface. This interface will enable the server designer to build the business scenario graphically, defining states and the transitions between them. The graphical interface will automatically translate that state chart into a business scenario made of WebSteps instructions, thus freeing the Web server designer from dealing with the WebSteps language syntax.

We also foresee the development of interacting business processes in order to accomplish more complex business tasks which involve several commercial actors.



page posting
welcome
if (search) goto (search criterias)
if (confirmation) goto (remind pre-reserv)
search
no results
form posting
search criteria
if (cancel) goto (welcome)
database query
if (results) goto (search results)
results
cancel
page posting
search results
if (presentation) goto (presentation)
if (new search) goto (search criteria)
confirmation
new search
back to list
presentation
new search
retrieve pre-reserv info
page posting
remind pre-reserv
if (reservation) goto (pers info)
else goto (welcome)
page posting
presentation
if (back to list) goto (search results)
if (pre-reserv) goto (pre-reservation)
if (new search) goto (search criteria)
incomplete: not all fields have been filled
reservation
cancel
pre-reserv
form posting
personal info
if (complete) goto (thank you)
form posting
pre-reservation
if (cancel) goto (presentation)
save pre-reserv info
if (reservation) goto (personal info)
else goto (confirm pre-reservation)
reservation
complete
insert DB personal info
send mail to owner
page posting
thank you
insert dB IDnumber
page posting
confirms pre-reservation

Figure 6: *State chart for the rent of apartments*

1.6 References

[Adobe 1997] Adobe Systems Incorporated. *Adobe PageMill. The Easiest Way to Create Pages for the World Wide Web.* http://www.adobe.com/Apps/PageMill/

[Berners-Lee et al. 1996] Berners-Lee, R. Fielding, H. Frystyk, *Hypertext Transfer Protocol HTTP/1.0* , RFC 1945, 1996

BEST COPY AVAILABLE

[Boutell 1997] Thomas Boutell, *MapEdit imagemap editing software*, http://www.boutell.com/mapedit/

[Diaz et al. 1995] Alicia Diaz, Thomas Isakowitz, Vanesa Maiorana and Gabriel Gilabert, RMC: A Tool to Design WWW Applications, *4th International WWW Conference, 11-14 December 1995*, Boston, MA, http://www.w3.org/pub/Conferences/www4/

[Fielding et al. 1997] R. Fielding, J. Gettys, J. Mogul, H. Frystyk and T. Berners-Lee, *Hypertext Transfer Protocol HTTP/1.1*, RFC 2068, January 1997

[Glintz 1995] M. Glintz, An Integrated Formal Model of Scenarios Based on State Charts, in Shaefer, Wad Botella, P. (eds), Software engineering-ESEC'95, *Proceedings of the 5th European Software Engineering Conference*, Springer Verlag, Lecture notes in Computer Science, No 989, pp. 254-271

[Kesseler 1995] Markus Kesseler, A Schema Based Approach to HTML Authoring, *4th International WWW Conference, 11-14 December 1995*, Boston, MA, http://www.w3.org/pub/WWW/Journal/1/kesseler.145/paper/145.html

[Kristol & Montuli 1997] Kristol, Montulli, RFC 2109 HTTP State Management Mechanism, February 1997

[Microsoft 1997] Microsoft Corporation, *Microsoft FrontPage Product Information*, http://www.microsoft.com/msoffice/frontpage/intro.htm

[Perl 1997] The Perl Language Home Page, http://www.perl.com/perl/

[Rumbaugh et al. 1991] J Rumbaugh Michael Blaha, William Premerlani, Frederick Eddy and William Lorensen, *State Charts of OMT: Object Oriented Modeling and Design*, Prentice Hall International Editions, 1991, pp. 123-137

Web Host Enumeration Through DNS

Domenico Dato, dato@di.unipi.it
Antonio Gullì, gulli@di.unipi.it
Giuseppe Attardi, attardi@di.unipi.it

SerRA (Servizi di Rete Ateneo)
Dipartimento di Informatica, Università di Pisa
Pisa, Italy

Abstract: Most search engines exploit spiders to implement the information gathering process. We present a technique to reduce the memory resources required by such spiders, improve the coverage and to provide for scaling through parallelization. The technique exploits the direct enumeration of Internet hosts through Domain Name Server. Tool bases on this technique have been successfully applied in a search engine for the ".it" domain but can be applied directly to other domains. We report statistics on the results of the use of our tools in this context.

World Wide Web and Robots

Web robots are tools used to gather data from the Web [De Bra 1996]. Robot behaviour may be formally described by means of an oriented graph $G = (N, A)$, where N is a finite set of nodes, corresponding to Web documents uniquely specified through their URL, and A is the set of arcs which represent unidirectional pointers, corresponding to the links between documents. G does not have to be fully interconnected since there can be URLs not reachable from other URLs: on Internet there can be a Web server completely autonomous defining a subgraph without any incoming arcs. A more appropriate model for the highly irregular World Wide Web is hence a set of connected components.

The task of a robot is to visit each node in G once avoiding to traverse twice the same path. This is done by building a sequence of graphs $G_1 \subset G_2 \subset \cdots \subset G_n \subseteq G$ each one representing a more precise approximation to the whole graph G.

The two main visit techniques are: the Depth-First-Search (DFS) and the Breadth-First-Search (BFS). With the first method, when a node is visited, we go away from it as much as we can until a blind alley is found (a node v all of whose adjacent nodes have already been visited). The order of visit is FIFO and the algorithm may, as a consequence, activate a recursive process whose depth is related to the length of the most extended not-cyclic path present on the Web. The second method, instead, visits the nodes according to the growing distance d from the start node r, where the distance between r and the generic node v is the length of the shortest path from r to v. The order of visit followed is, in this case, LIFO.

Often the two techniques are combined: a set of starting points for the search is built by performing a BFS with maximum distance from a root, and then DFS is applied from this set. All the other methods are variants of combinations of BFS and DFS.

Note that the whole graph G is not known a priori. New nodes are added each time an already acquired arc is followed, building a new graph G_i from the given G_{i-1}.

An analysis of the above algorithms leads us to reveal some limits:

> *Paths Memory*: to guarantee the correctness of the visit algorithm, an auxiliary structure must be used (generally a database) to hold every already visited node (every already reached URL). The maintenance of this structure is particularly time consuming because the quantity of documents actually available on the Web is considered to be approximately 40 millions pages.
>
> *Low scalability*: the Web indexing process is well suited to be executed in a parallel way by mean of a prefixed number R of robots that follow the network on different parts of the graph G used as a model. In De Bra [DeBra 1996] calculates that, given a starting base of twenty millions Web documents (containing about a thousand gigabytes of text) and with an average transfer rate of 3 Kbyte/s, the time that must be used to fetch all the information with a single robot is equal to about 8 months. The same task, realized trough ten distinct robots, may be accomplished in about 20 days. By adopting a parallel indexing architecture, a new class of problems is introduced. First at all, because of the fact that the structure isn't known a priori, the structure in which the URL are stored must be distributed among the robots. Every visited document is stored with a write disk access; every time a document is fetched a check must be performed, by using a read disk access, to be sure that this hasn't already been done by another robot. More robots are present, more probable is the probability of use conflicts.
>
> *Balance Absence*: in order to achieve good performance levels, an architecture based on a high robot parallelization level must use a balance policy for assigning work.
>
> *Root identification*: since G is not fully connected, a full search requires selecting at least one root node within each connected component. Connectivity within each component is low, as demonstrated by statistical survey performed within the RBSE project [Eichmann 1994] showing that 59% of Web documents have only one link and 96% less then five links. Therefore the choice of the root within each component it is quite critical.

We present some tools we developed for enumerating hosts and facilitating Web search. These tools are used in ARIANNA[1], an Internet search engine for Web sites with italian language content. Since some hosts with italian language content do not appear under the ".it" DNS domain, several heuristics, whose discussion is beyond the scope of this paper, have been employed in order to find them.

Direct WWW Enumeration

The method we implemented exploits the fact that many problems that affect the Web visit algorithms, may be solved by building a priori an URLs list that must be the widest possible. Each item is used as a starting point for a distinct robot. Similar techniques have already been proposed and a list of them is presented in [DeBra 1996]. However, most of these are ad hoc and not general enough. For example one suggestion is to survey Usenet News postings in order to discover URLs. An important aspect of the method we propose is to be parametric: different selection criteria give rise to a spectrum of solutions. A very selective criterion, even though it reduces the indexing time, could discard valid hosts; a weak selective criterion, on the other hand, could show the opposite behaviour.

---

[1] http://www.arianna.it or http://www.arianna.com

A hierarchical selection criterion based on DNS name

The WWW structure is inherently chaotic. The final aim of BFS and DFS algorithms is to build a spanning tree for the connected components of the Web. A spanning tree suggests the idea to find a way of hierarchical ordering net hosts. Luckily, this ordering can be derived from the domain subdivision provided by the DNS system [Albiz & Liu 1996]. Using tools like *host* or *nslookup* and directly querying authoritative nameservers one can find the set of host names $H(D_i) = \{h_{d1}, \ldots, h_{di}\}$ inside the domain D_i. Let $D = \{D_1, \ldots, D_z\}$ be the set containing all Internet domains and let H be set of all Internet hosts defined as follows:

$$H = \bigcup_{D_i \in D} H(D_i)$$

Let $n = \#(H)$. R robots (with $R << n$) can be used to examine the n found hosts in parallel using DFS or BFS techniques (or their variants). Since not all Internet hosts have a WWW server, assigning to a robot a machine that contains nothing to be indexed is a waste of time and could invalidate any balancing policy. The compromise we adopted is to preselect all Internet hosts having a canonical name (record A) or an alias (CNAME) among those most commonly used for machines running WWW services. In detail, we define a criterion K according to which a host becomes candidate to be indexed: for instance we can test whether its name contains "www", "web", "w3" as substrings. Let W be the set of all hosts satisfying K:

$$W = \{h \in H \mid h \text{ satisfies } K\}$$

Let $Q = (P_1, \ldots, P_R)$ be a partition of W: the indexing domain $DI(r)$, assigned to a robot (with $1 \quad r \quad R$) is defined as $P_r \in Q$.

Our method has the following properties:

path memory of robot r is limited by the number of Web documents contained in Web servers indexed by r.

good scalability with respect to the number of used robots. Since $P_1, \ldots, P_R$ are disjoint there is no need to keep shared structures to record the paths already followed by each robot.

the choice of different substrings (new definition of K criterion) allows us to modify the cardinality of W.

A further optimization allows avoiding direct query to an authoritative nameserver for each domain in order to build the set H. To reach this aim the RIPE[2] monthly survey (*hostcount*) is exploited. This has the advantage of saving bandwidth: one avoids building the Italian domain topology since it is already available on Internet. [Tab. 1] reports the results obtained applying criterion K on the *hostcount* for the italian domain. However, this improvement has introduced two problems. Let H_{RIPE} be the host used by RIPE to survey the net and let $NS(D) = \{NSP(D), NSS_1(D), \ldots, NSS_j(D)\}$ be the set formed by the primary nameserver and by j secondary nameservers for domain D. It can happen that:

a query made from H_{RIPE} to all nameservers in $NS(D)$ is not answered, for instance because of a time-out.

the policy of a domain forbids zone retrieval from H_{RIPE}.

If one of these conditions occurs during analysis of domain D_i, *hostcount* omits from the survey both D_i and all domains delegated by the authoritative nameservers for D_i. Note that (1) and (2) might not happen if the queries to $NS(D)$ are performed from a host different from H_{RIPE}. In such cases, one could discover the domain structure even if it is not present in *hostcount*. Statistical studies (reported in [Tab. 2]) justify this corrective intervention: in fact for domain ".it" our method counts a number of machines that is more than 10% greater than that contained in the RIPE database.

---

[2] Réseaux IP Européens, http://www.ripe.net

Once the set of the hosts in W has been built, in order to determine those which are actual Web servers, we have developed *testwww*, an agent which tests the presence of a HTTP server on a set of TCP ports chosen among those generally used to provide such service. *testwww* was tested on more than 14000 Italian hosts [Tab. 3]. Let T be the set built from W using *testwww*:

$$T = \{t \mid t = testwww(w), w \in W\}$$

We then partition T into R subsets ($P_1, \ldots, P_R$) as above and assign a robot r to each $P_i = DI(r)$. During the analysis of host h $DI(r)$, the path memory for robot r is limited to the documents contained in h. This provides scalability for performing parallel indexing in a search engine.

Besides, *testwww* allows us to build statistics on [Tab. 4]:

- the choice of ports used for the WWW service;
- the kind of Web server implementation;
- other services offered beyond WWW (such as Proxy-Server).

Document Relocation

Some HTTPD daemons provide several virtual Web servers on the same IP address. This is known as VIRTUAL HOSTING: the administration job is centralized while the content of each virtual site is under the responsibility of its respective owner. A single IP address is used but each virtual site has assigned a CNAME (or an A record on the same IP). Virtual Hosting was standardized in version 1.1 of HTTP protocol [Berners-Lee, Fielding & Frystyk 1996]. This requires a revision of robot's visiting algorithms. Indeed, if path memory is managed only according already visited IP addresses, documents with the same document path but located on distinct virtual servers will be missed.

We solved this problem using a heuristic: the *testwww* agent, given a set of A and CNAME records $N(host) = \{ A_1, \ldots, A_p, C_1, \ldots, C_q \}$ associated via DNS to net address I_{host}, contacts $(p + q)$ times the HTTP server on I_{host} asking for the Document Root, with a different argument for the "*Host:*" directive chosen from $N(host)$. For each returned page, *testwww* computes a MD5 RSA [Schneier 1996] signature. Two Web servers are considered distinct if and only if they have different associated signatures. This is well represented by the set M:

$$M = \{m \mid m = MD5(n), \forall n \in N(host), \forall host \in T\}$$

Robots' Load Balancing

To perform the partitioning of hosts into indexing domains we decided to exploit the division into Autonomous Systems (AS). The idea is to minimize the number of border gateways (according to BGP4 terminology [Huitema 1995]) crossed during the indexing phase in order to reduce the probability of using links with high traffic load.

For the case of the italian domain we chose three indexing hosts, each one in a different AS. The *prtraceroute* tool [PRIDE 1996] tells us to for each host, both which AS it belongs to and how many AS are crossed to reach it. We then filter static information given by the RIPE Registration Authority (via *whois*) in order to assign hosts to the indexing points so that the number of crossed border gateways is minimized and therefore the indexing time is reduced.

Open Research Areas

In summary our method has the following features:

reduced indexing time since each robot has a preassigned task to be accomplished;

reduced use of resources because there is no need to maintain shared structures;

potential for parallelism in terms of robots that can be used;

possibility of assigning work according to a policy of load balancing.

As it may be expected there also are some limits: the most evident one is that we miss Web servers not satisfying the K criterion. Several approaches are possible to mitigate this drawback although we expect it not to be significant. Statistical data (reported in [Tab. 5]) show that we collect a remarkable number of URLs compared to similar engines based on traditional spider mechanisms: ARIANNA is currently the most complete Italian search engine in terms of indexed information.

A first method to discover a host not satisfying the criterion K is based on a different way of integrating DNS enumeration techniques with visit algorithms. Once an indexing domain $DI(r) = \{ H_1, \dots, H_l \}$ has been built, each robot is allowed to cross not only the host H_i (with $i < l$) but also all hosts in the same DNS domain. In this case the path memory of each robot is now limited to WWW pages contained in the Web servers of the examined DNS domain. Careful partition that doesn't assign to different robots the same DNS domain allows us to reach scalability for parallel indexing. Alternatively the method may use as a boundary the IP classes.

A second method is based on post processing the already collected Web pages when the indexing phase is ended. This is done in order to discover news URLs not enumerated in the set W. Note that this is a local job (not involving communication) and as consequence less time-consuming.

Experimental Results

We report statistical data which refer to the use of our tools within the ARIANNA search engine.

Criterion applied on *hostcount*	Month	Res.	Month	Res.
Number of tested domain	Dec 96	6995	Feb 97	8608
WWW servers discovered	Dec 96	6932	Feb 97	8573

Tab 1: Criterion applied on *hostcount*

hostcount Problems	Month	Res .	Month	Res.
Unanswered queries	Dec 96	646	Feb 97	464
AXFR negations	Dec 96	248	Feb 97	364
Wrong answers from NS	Dec 96	484	Feb 97	401

Tab 2: *hostcount* Problems

Direct DNS query	Month	Res.	Month	Res.
Tested domains	Dec 96	1231	Feb 97	738
WWW servers discovered	Dec 96	513	Feb 97	134

Tab 3: Direct DNS query

testwww	Month	Res.	Month	Res.
Examined domains	Dec 96	8228	Jul 97	14431
Examined hosts	Dec 96	7440	Jul 97	13780
Hosts with an active Web server	Dec 96	6741	Jul 97	11045
Proxy servers discovered	Dec 96	424	Jul 97	n.a.

Tab 4: *testwww* Results

ARIANNA	Month	Res.	Month	Res.
Total number of indexed sites	Dec 96	6.629	Jul 97	8.134
Total number of reached URLs	Dec 96	1.351.442	Jul 97	2.095.318
Robots' disk space used (KB)[3]	Dec 96	10.151.481	Jul 97	7.558.919
Information Retrieval's disk space (KB)	Dec 96	6.932.135	Jul 97	9.879.552
Max object's num for site	Dec 96	17.187	Jul 97	127.173
Average time for site	Dec 96	2 hours	Jul 97	2 hours
Total time ARIANNA	Dec 96	15 days	Jul 97	20 days

Tab 5: ARIANNA Results (thanks to OTM Interactive Telemedia Labs for these data)

Acknowledgements

The work described here was done within the cooperation between SerRA, Università di Pisa and OTM Interactive Telemedia Labs. We would like to thank A. Converti, L. Madella, A. Vaccari and G. Antichi of the OTM labs for their support.

References

[Albiz and Liu 1996] Albitz, P., & Liu, C. (1996). DNS and BIND 2nd Edition. O'Reilly & Associates.

[PRIDE 1996] PRIDE - Policy based Routing Implementation Deployment in Europe, ftp://ftp.ripe.net/pride/

[DeBra 1996] De Bra, P.M.E (1996). Introduction to WWW Robot Technology, 5th International World Wide Web Conference. O'Reilly & Associates.

[Eichmann 1994] Eichmann, D. (1994). The RBSE Spider-Balancing Effective Search Against Web Load, First WWW Conference.

[Berners-Lee, Fielding and Frystyk 1996] Berners-Lee T., Fielding R., Frystyk H. (1996). Hypertext Transfer Protocol – HTTP/1.1 – RFC 2068.

[Huitema 1995] Hiutema, C. (1995). Routing In The Internet. Prentice Hall.

[Schneier 1996] Schneier, B. (1996). Applied Cryptography, John Wiley, 2nd edition.

---

[3] The space decreased due to an optimization in the search program

Creating Adaptive Hyperdocuments for and on the Web

Paul De Bra[*]
Department of Computing Science
Eindhoven University of Technology
PO Box 513
NL 5600 MB Eindhoven
The Netherlands
debra@win.tue.nl

Licia Calvi
Department of Romanic Languages
University of Antwerp
Universiteitsplein 1
B 2610 Wilrijk
Belgium
calvi@uia.ua.ac.be

[*]Paul De Bra is also affiliated with the University of Antwerp, and with the "Centrum voor Wiskunde en Informatica" in Amsterdam.

Abstract: The course "2L690: Hypermedia Structures and Systems" is taught entirely through World Wide Web, and offered at six different universities in the Netherlands and Belgium. Different approaches have been taken towards adding adaptivity to the course text. This paper reviews these development steps and presents the final design, which results in adaptive hyperdocuments that can be written in standard HTML 3.2, possibly by using off the shelf HTML editors. We also present a simple but powerful representation of user (student) knowledge, as used to adapt the link structure and textual contents of the course text.

Keywords: adaptive hypertext, courseware, knowledge representation, standard HTML

1. Introduction

Many different definitions of and techniques for adaptive hypermedia exist. A good overview is presented in [Brusilovsky 1996]. In this paper we follow the terminology introduced in that overview to characterize the kinds of adaptation introduced in the course "2L690: Hypermedia Structures and Systems". We describe not only the adaptation techniques used for this course, but also compare them with some other initiatives for using adaptive hypertext in courseware, such as a C Programming course [Kay & Kummerfeld 1994a, Kay & Kummerfeld 1994b] and the ELM-ART Lisp course [Brusilovsky et al 1996a].

World Wide Web was not designed with highly dynamic applications in mind. A typical characteristic of adaptive hypertext is that during the reading process the presentation of an information item (e.g. a page) may be different each time that item is revisited. The way some WWW browsers deal with their history mechanism makes it difficult to ensure that pages are reloaded each time they are (virtually) modified on the server. This problem can be resolved with some browsers (like Netscape Navigator) but on some others (like NCSA Mosaic for X) it cannot. Unfortunately the browsers cannot be blamed for their behavior because the way they implement their history mechanism satisfies the requirements set out by the HyperText Transfer Protocol (HTTP) standard.

Authoring adaptive hypermedia is also difficult in a WWW environment, because the HyperText

Markup Language (HTML) has no provision for "conditional text". It is not possible to create a single HTML document of which only selected parts are presented to the user, based on some kind of environment variables controlled by an agent that monitors the user's knowledge state or preferences. The Interbook tool [Brusilovsky et al.\ 1996b] for instance uses *concept-based indexing* to provide access to (non-adaptive) HTML pages from dynamically generated index pages. Displaying index and "content" pages simultaneously is done through *frames*, a technique introduced by Netscape but which is not available in all browsers, and which is not part of the latest HTML-3.2 standard.

Brusilovsky [Brusilovsky 1996] distinguishes two main categories of adaptivity:

- **Adaptive presentation**: these techniques consist of both the selection of different media depending on user preferences and the adaptation of a document's textual (or multimedia) content based on a user's knowledge state.
- **Adaptive navigation support**: these techniques change the (apparent or effective) link-structure between the pages that together make up a hyperdocument. Brusilovsky distinguishes between *direct guidance* (suggest a next best link), *adaptive sorting of links* (displaying a list of links from best to worst), *adaptive hiding of links* (hiding inappropriate links), *adaptive annotation of links* (marking links as appropriate or inappropriate but not deleting any) or *map adaptation* (changing graphical overviews of the link structure).

The courseware for "2L690: Hypermedia Structures and Systems" contains both forms of adaptivity. The textual contents of pages is adapted to the knowledge state of the student. This paper describes a new way of encoding conditional text in (standard) HTML documents. This new way supersedes a first attempt at providing adaptive content, described in [De Bra & Calvi 1997]. The course offers adaptive navigation support by means of adaptive hiding of links. Only links to pages which are "interesting" for the student to read next are shown. The technique used for this navigation support is described in [Calvi & De Bra 1997].

We propose a simple authoring environment for adaptive hypertext courseware, which offers the following features:

- The content of the pages of the course text is adaptive, as well as the link structure.
- The adaptive documents are written in standard HTML, and can be authored using (some of the existing) HTML editors.
- Pages of the course text, as well as tests and assignments that may be embedded in the course text, generate knowledge about *concepts*.
- Information items (ranging from words to large parts of documents) and links can be made dependent on Boolean combinations of concepts (using *and*, *or*, *not* and arbitrary parentheses).
- A verification tool lets authors check whether all the information in the course can be reached by a student, independent of the order in which the student decides to view pages. This problem is non-trivial since knowledge about certain concepts can make

pages inaccessible.

2. Why Adaptive Hypertext on World Wide Web is Difficult

World Wide Web was not designed with highly dynamic applications in mind. This is not simply a matter of oversight in the standards definitions, but more of the programmers and companies who first created WWW-browsers and servers, as well as of document authors who only used a fraction of what the HTML and HTTP standards have to offer. Here are a few examples of problems with the current standards and practice:

- In an adaptive hypermedia system following a number of links forwards and then backtracking may result in changes to the previously visited documents. The HTTP standard does not require backtracking to request the documents from the server again, not even when the documents are expired. This means that there is no guaranteed way for a server to tell the browser to reload a document when the user uses the "back" button to revisit that page.

  > *By default, the Expires field does not apply to history mechanisms. If the entity is still in storage, a history mechanism should display it even if the entity has expired, unless the user has specifically configured the agent to refresh expired history documents.*
  > (quoted from RFC's 1945 and 2068, which define HTTP/1.0 resp. 1.1)

  The HTTP standard acknowledges that some browsers may let users configure the history mechanism to verify whether a document is modified even when going through the history mechanism. Unfortunately the standard does not *require* browsers to have such a feature, and specifies that the default behavior should be *not* to verify whether the document has changed.

- Although HTTP offers at least the possibility to suggest that pages should be reloaded by declaring them to be expired, many authors of HTML documents never indicate that a document may expire at some given date, not even when they know exactly when a new version will replace the current one. A possible reason for this is that HTML (either version 2.0 as defined by RFC 1866 or the newer version 3.2) does not offer an <EXPIRES> tag to indicate an expiry date. Authors have to use the <META> tag to force the server to generate an HTTP Expires field, using the following syntax:

  ```
  <META HTTP-EQUIV="Expires"
     CONTENT="Tue, 04 Dec 1993 21:29:02 GMT">
  ```

  The HTTP/1.0 standard encourages the expiry mechanism to the point that an invalid date of 0 should be interpreted as "expires immediately". (Sadly, this "encouragement" has been dropped in the HTTP/1.1 standard.) Still, expiry has not yet become sufficiently popular to warrant that all browsers interpret expire-fields correctly. Furthermore, apart from browsers there are also a number of proxy-caches that do not yet understand the HTTP/1.1 caching directives that have been introduced to avoid caching expired or rapidly changing documents.

  > *Note: Applications are encouraged to be tolerant of bad or misinformed*

implementations of the Expires header. A value of zero (0) or an invalid date format should be considered equivalent to an "expires immediately." Although these values are not legitimate for HTTP/1.0, a robust implementation is always desirable.
(quoted from RFC 1945 which defines HTTP/1.0)

- HTML does not offer a possibility to conditionally include text or multimedia objects. There is no such thing as an "<IF>" tag.
    - Several attempts have been made to use the Unix C preprocessor for conditional pieces of content, but mixing C preprocessor commands (like **`#ifdef`** clauses) with HTML encoded documents generates source text that is difficult to write and read. The previous edition of the course "Hypermedia Structures and Systems" (with code 2L670) used a mix of **`#if`** and **`#ifdef`** constructs to achieve adaptive content [Calvi & De Bra 1997]. The newer edition, course 2L690 which started in the fall trimester of 1997, uses the approach described in this paper.
    - Pim Lemmens [Lemmens 1996] has proposed and implemented a mechanism for parameterizing Web pages, by means of &...; constructs. This suggestion works well for small textual variations, like including **`&date;`** in a document to generate the current date, or to offer alternative wordings for technical terms. Referring to a page as **`<A HREF="test.html?node=page>`** would result in all occurrences of **`&node;`** being replaced by the word "page", while **`<A HREF="test.html?node=node>`** would generate a document in which every **`&node;`** is displayed as "node" (which would be done after explaining what a "node" is). Although this construct looks like valid HTML it is not, and cannot be generated through a strict HTML editor.
    - In HTML tags are case insensitive. A "smart" preprocessor can therefore interpret tags written in lowercase differently from tags in uppercase. Course 2L690 uses this possibility to distinguish between *conditional links* (authored as **`<a href=...>`**) and *unconditional links* (authored as **`<a HREF=...>`**) [De Bra & Calvi 1997].
- An interesting and promising possibility is offered by scripting languages such as JavaScript (developed by Netscape Communications). Using JavaScript one can embed different variations of a document's content in a single file, and make the browser present the appropriate elements based on the values of some variables that can be generated by the agent which monitors a user's knowledge state or preferences. Unfortunately JavaScript is still heavily under development. Only a few browsers offer scripting and their definitions and implementations of JavaScript are incompatibly different. The HTML-3.2 definition is still only partially "script-aware", meaning that a **`<SCRIPT>`** tag has been defined as a placeholder, but the current JavaScript practice to include method calls in anchor and button tags is not (yet) allowed.

3. Encoding Knowledge and Conditional Text in HTML

Hypertext techniques and World Wide Web technology have been used in educational settings, mostly for computer-science courses where students have to master certain skills such as programming (in C [Kay & Kummerfeld 1994a, Kay & Kummerfeld 1994b], or in Lisp [Brusilovsky et al. 1996a]). The hypertext (link) structure in such courses is fairly simple, since learning a programming language is a mostly linear process. Indicating which chapters are still to be avoided and which pages to read first is easy.

In the course "2L690: Hypermedia Structures and Systems" the link structure is made complex on purpose: students learn about the concepts of hypertext, and the best way to do so is by experiencing hypertext. Not all link structures are equally easy to navigate through. In course 2L690 the "chapters" which appear to exist when the student looks at the first page are actually overlapping sets of pages. For a number of information nodes it is impossible to tell to which (unique) chapter they belong. The course contains some introductory chapters, giving definitions and a historical overview, and advanced chapters, describing reference models, navigation and retrieval problems, authoring issues and multi-user aspects. It is desirable for students to first read the introductory chapters, and therefore to advise or force them to do so (by dimming, hiding or removing links to the advanced chapters at first). Nonetheless an introductory chapter and an advanced chapter may share common pages. This makes enabling or disabling access to pages more complicated than simply enabling or disabling whole chapters, and it may also suggest using "simple" wording when a page is read as part of an introductory chapter, and more technical wording when that same page is read as part of an advanced chapter.

In order to monitor the student's progress and knowledge state *concepts* are associated with pages from the course text. Each concept is denoted by means of a single word. (Multiple words can be simulated by joining them using underscores instead of spaces.) Much like in [Rosis et al. 1994] the concepts are collected in a *Dictionary of Concepts.* While for programming language and similar courses the user-model needs to consist of both "KNOW-ABOUT" and "PRACTICE-IN" facts, we currently make no such distinction. The knowledge state of a student is simply a set of concepts the student has read about (or successfully taken a test about).

For each page of the course text a number of concepts may be prerequisite knowledge, and a number of (other) concepts may make the page superfluous. In [De Bra & Calvi 1997] this prerequisite and/or forbidden knowledge is used to determine whether to enable or disable links to a page. More complex Boolean combinations were not possible in that proposal.

Depending on the knowledge state of the student not only links to pages but also the contents of pages may need to be adapted. In [Calvi & De Bra 1997] we proposed to use C-preprocessor (**`#ifdef`**) constructs to achieve this goal. However, mixing HTML with C-preprocessor statements makes authoring unnecessarily complicated. Since we aim to provide an authoring environment which is also suited for the development of non computer related courses, authoring needs to be simple and intuitive.

In our new proposal both whole pages, links to pages and pieces of HTML text (possibly including images), can be enabled or hidden depending on a Boolean combination of concepts. We use HTML comments to mix "if-statements" with HTML text, and use a filter to select the appropriate parts of the HTML page and sent those to the user's browser. The following example of source text shows what the adaptive index-page for the hypermedia course text looks like:

```
<!-- requires true -->
<!-- generates index -->

  <H1>Hypermedia structures and systems</H1>
  Welcome to course 2L690 at the Eindhoven University of Technology.
  <P>

<!-- if not readme -->

  Since you are just beginning to browse through this course,
  you should first read
  <a href="readme.html">the instructions</A>.
  These will explain how to use this course text,
  together with a graphical World Wide Web browser such as the
  Netscape Navigator, Microsoft Internet Explorer, or NCSA Mosaic.
  In order to get to the instructions you must
  click (the left or only mouse button) on the phrase
  <a href="readme.html">"the instructions"</a>.<BR>
  <img src="caution.gif" align="bottom"> You cannot start
  reading the dynamic course text until you have read these
instructions.
  <p>
  The items below indicate (not necessarily disjoint) parts of the
course
  text, which will become accessible after you have read
  <a href="readme.html">the instructions</a>.

<!-- else -->

  This course contains the following (not necessarily disjoint) parts:

<!-- endif -->

  <ul>
  <li><a href="intro.html">Introduction</a>
    (it is advised to read this before the other items)
  <li><a href="definition.html">Definition of hypertext and
hypermedia</a>
  <li>The <a href="history.html">history</a> of hypertext and
hypermedia

<!-- if readme but not (introduction and definition and history) -->

  </ul>
  The following parts will become available later (when you are ready
for them):
  <ul>

<!-- endif -->

  <li>The <a href="architecture.html">architecture</a> of hypertext
systems
```

```
  <li><a href="navigation.html">Navigation</a> (and browsing semantics)
in hypertext
  <li><a href="retrieval.html">Information Retrieval</a> using
hypertext
  <li><a href="authoring.html">Writing</a> hypertext
  <li><a href="distribution.html">Distribution and Concurrency</a>
issues
  <li><a href="future.html">The Future of Hypertext and Hypermedia</a>
  <li><a href="assignment.html">Assignment for this course</a>
  </ul>
```

Each page of the course text starts with a comment that indicates which Boolean combination of concepts is *required* to allow access to the page. (In the example, "true" means that nothing is required.) The second comment indicates which concept(s) are *generated* by visiting the page. The latter concepts are added to the student's knowledge **after** reading the page. This implies that text fragments that depend on the concepts generated by a page are not displayed the first time the page is visited. It also implies that a page may forbid the same concept(s) it generates, in which case it will be accessible only once. Section 4 explains that one needs to be careful with the selected combinations of required and generated knowledge. Should a page require a concept it generates, the page will never be accessible unless another accessible page generates the same concept.

Note that although the links to all chapters are always present in the source text, the software described in [De Bra & Calvi 1997] will hide (remove) these links until the student has gained sufficient knowledge. We could have used more "`<!-- if -->`" commands (actually comments in HTML) to include these links conditionally, but that would make the document source much harder to write (and read). The software of [De Bra & Calvi 1997] is kept in place because it is also needed for maintaining each student's individual log file. (The use of log files is described in [De Bra 1996].)

Note also that the keyword *but* in the second "`<!-- if -->`" statement is simply used as a synonym for *and*, but is closer to natural language.

When the student first looks at this page, the following text will be presented:

Hypermedia structures and systems

Welcome to course 2L690 at the Eindhoven University of Technology.

Since you are just beginning to browse through this course, you should first read the instructions. These will explain how to use this course text, together with a graphical World Wide Web browser such as the Netscape Navigator, Microsoft Internet Explorer, or NCSA Mosaic. In order to get to the instructions you must click (the left or only mouse button) on the phrase "the instructions".

⚠You cannot start reading the dynamic course text until you have read these instructions.

The items below indicate (not necessarily disjoint) parts of the course text, which will become accessible after you have read the instructions.

- Introduction (it is advised to read this before the other items)
- Definition of hypertext and hypermedia
- The history of hypertext and hypermedia
- The architecture of hypertext systems
- Navigation (and browsing semantics) in hypertext
- Information Retrieval using hypertext
- Writing hypertext
- Distribution and Concurrency issues
- The Future of Hypertext and Hypermedia
- Assignment for this course

After reading "the instructions" the page looks like:

Hypermedia structures and systems

Welcome to course 2L690 at the Eindhoven University of Technology.

This course contains the following (not necessarily disjoint) parts:

- Introduction (it is advised to read this before the other items)
- Definition of hypertext and hypermedia
- The history of hypertext and hypermedia

The following parts will become available later (when you are ready for them):

- The architecture of hypertext systems
- Navigation (and browsing semantics) in hypertext
- Information Retrieval using hypertext
- Writing hypertext

- Distribution and Concurrency issues
- The Future of Hypertext and Hypermedia
- Assignment for this course

Finally, after also reading (some pages of) the first three chapters the presentation changes to:

Hypermedia structures and systems

Welcome to course 2L690 at the Eindhoven University of Technology.

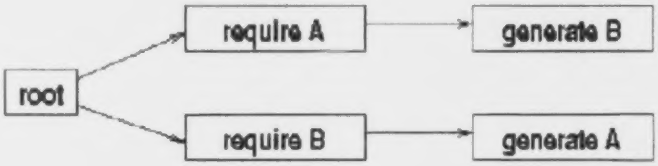
This course contains the following (not necessarily disjoint) parts:

- Introduction (it is advised to read this before the other items)
- Definition of hypertext and hypermedia
- The history of hypertext and hypermedia
- The architecture of hypertext systems
- Navigation (and browsing semantics) in hypertext
- Information Retrieval using hypertext
- Writing hypertext
- Distribution and Concurrency issues
- The Future of Hypertext and Hypermedia
- Assignment for this course

Conditional content need not necessarily be tied to "real" knowledge gained by the user. In the courseware for 2L690 the user can manually set knowledge on or off through a setup page. By switching knowledge of a **verbose** "concept" on or off one can give the user the option of selecting or deselecting optional additional content. It is thus possible to give the user a choice between different presentations of the same course text.

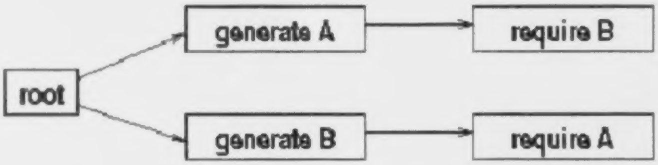
4. Validating Adaptive Hypertext Link Structures

In a static hypertext analyzing whether all nodes can be reached (from a given "root" node) is a matter of simple graph traversal. The Boolean conditions however complicate the link structure, because links may or may not be present depending on the student's knowledge state. As long as no negation (*not* operator) is used the *reachability problem* is still easy: one can repeat the process of trying to follow links (forwards) each time some knowledge is gained. The following figure illustrates a case of unreachable pages:



This figure illustrates that no pages other than root can be accessed because the pages that generate concepts cannot be reached without passing through pages that require that knowledge.

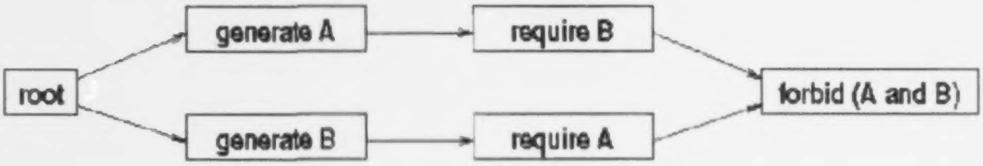
Reachability is not a sufficient condition for *easy* readability (actually "navigability"). The following figure illustrates an undesirable construct:



In order to reach the nodes that require knowledge about A and B the student must visit the node that generates A, then go back to the root, then visit the node that generates B, then the node that requires A, then go back to the root, revisit the node that generates A, and then visit the node that requires B. One can easily prove that the student must revisit a page by following links forwards. An "easy" link structure does not require the student to do this.

Of course a student may always find ways to navigate through a course text that do require revisiting a page several times. A quality measure for link structures is not the absence of such paths, but the presence of navigation paths that contain no required page revisits (except by backtracking).

The presence of negation complicates matters even further. Gaining knowledge may make (links to) pages unavailable. In the example of the index page for the hypermedia course, the link to the "instructions" page disappears after reading the instructions. (The page can still be reached through a complete index, as described in [De Bra 1996].) In this case the disappearing link does not make reading a page impossible because the link deletion happens *after* the student has read the instructions. However, in general it is possible that by reading pages in a peculiar order some pages may never become available to the student. The quality measure for link structure is the absence of navigation paths that leave some pages inaccessible at all times. The figure below illustrates how negation may make a page inaccessible in a more subtle way:



One can easily see that the page that depends on not knowing both concepts A and B can never be reached.

We are currently building a simple tool that verifies the link consistency of a course text. In particular the tool analyzes the following properties:

- Are there navigation paths that make it impossible to visit some page(s)? If so, which pages may not be reachable? (This can be a consequence of using negation, but also of requiring concepts which are never generated before they are needed.)
- Are there (conditional) parts of pages that can never be viewed (no matter which navigation path is used)?
- Is it possible to navigate through the whole course text without ever following a forward link to a page that was visited before? (Given a course text like 2L690 this implies that it must be possible to read the text chapter by chapter.) If not, which pages must be revisited in order to gain access to which pages?

5. Conclusions and Future Work

Creating adaptive hypermedia documents, that have a complex (non-hierarchical) structure, is difficult in general. Analysis tools may be needed to help authors verify that their adaptive hyperdocuments are easy to navigate through. We are currently building such a tool that will be used not only for the next version of course "2L690: Hypermedia structures and systems", but also for a course on Italian economy and a course on Graphical User-Interfaces. Course 2L690 is updated about every 6 months. The first version with adaptive content was installed in January 1997. The version using the technology described in this paper has become operational the fall of 1997. A student is currently investigating how the adaptive linking introduced in the fall of 1996 [De Bra & Calvi 1997] has influenced the browsing behavior of student, as compared to the previous (non-adaptive) version described in [De Bra 1996]. Informal interviews with students have already confirmed that adaptive linking alone is insufficient, because hiding links without any additional explanation (which could be conditionally included) is frustrating for the reader.

Our first attempt to use the (Unix) C-preprocessor for the creation of adaptive content resulted in an awkward authoring environment in which two completely different syntaxes had to be mixed. This resulted in source texts that were difficult to write and read. The approach proposed in this paper uses standard HTML, which enables authors to use HTML editors (or generators) for writing adaptive hyperdocuments for and on the Web. Besides conditional constructs in HTML we rely on the software described in [De Bra & Calvi 1997] for conditionally hiding links. This significantly reduces the number of conditionals authors have to include in the source text of their documents. All the (current) software is written in Java. For performance reasons the Web server for the courseware had to be upgraded from a 486-66 to a Pentium-Pro 200 (both running

Solaris x86 2.5.1).

6. References

[De Bra 1996]

De Bra, P., Teaching Hypertext and Hypermedia through the Web. In: Proceedings of the WebNet'96 Conference, pp. 130-135, San Francisco, 1996.
(URL:http://wwwis.win.tue.nl/~debra/webnet96/index.html)

[De Bra & Calvi 1997]

De Bra, P., Calvi, L., Improving the Usability of Hypertext Courseware through Adaptive Linking. Proceedings of the Flexible Hypertext Workshop, Southampton, 1997.
(URL: http://wwwis.win.tue.nl/~debra/flex97/)

[Brusilovsky 1996]

Brusilovsky, P., Methods and Techniques of Adaptive Hypermedia. In: User Modeling and User-Adapted Interaction **6**: 87-129, Kluwer academic publishers, 1996.
(URL: http://www.contrib.andrew.cmu.edu/~plb/UMUAI.ps)

[Brusilovsky et al. 1996a]

Brusilovsky, P., Schwarz, E., Weber, G., ELM-ART: An intelligent tutoring system on World Wide Web. Proceedings of the Third International Conference on Intelligent Tutoring Systems, ITS-96, Montreal, 1996. (Lecture Notes in Computing Science, vol. 1086, pp. 261--269).
(URL: http://www.contrib.andrew.cmu.edu/~plb/ITS96.html)

[Brusilovsky et al. 1996b]

Brusilovsky, P., Schwarz, E., Weber, G., A Tool for Developing Adaptive Electronic Textbooks on WWW. Proceedings of the WebNet'96 conference, pp. 64-69, San Francisco, 1996.
(URL: http://www.contrib.andrew.cmu.edu/~plb/WebNet96.html)

[Calvi & De Bra 1997]

Calvi, L., De Bra, P., Using Dynamic Hypertext to create Multi-Purpose Textbooks. (to appear) In: Proceedings of ED-MEDIA'97, Calgary, 1997.
(URL: http://wwwis.win.tue.nl/~debra/ed-media97/)

[Kay & Kummerfeld 1094a]

Kay, J., Kummerfeld, B., An Individualised Course for the C Programming Language. In: Proceedings of the Second International WWW Conference "Mosaic and the Web", Chicago, 1994.
(URL: http://www.ncsa.uiuc.edu/SDG/IT94/Proceedings/Educ/kummerfeld/kummerfeld.html)

[Kay & Kummerfeld 1994b]

Kay, J., Kummerfeld, B., Adaptive Hypertext for Individualised Instruction. Workshop on Adaptive Hypertext and Hypermedia at the Fourth International Conference on User Modeling, 1994.
(URL: http://www.cs.bgsu.edu/hypertext/adaptive/Kay.html)

[Lemmens 1996]

Lemmens, W.J.M., Use of Parameterised Hypertext Pages. Eindhoven Univ. of Technology, 1996.
(URL: http://wwwis.win.tue.nl/~wsinpim/parmpage.html)

[Rosis et al. 1994]

de Rosis, F., De Carolis, B., Pizzutilo, S., User-Tailored Hypermedia Explanations. Workshop on Adaptive Hypertext and Hypermedia at the Fourth International Conference on User Modeling, 1994.
(URL: http://www.cs.bgsu.edu/hypertext/adaptive/deRosis.html)

Virtually Deschooling Society:
Authentic Collaborative Learning via the Internet

R. T. Jim Eales
Department of Computer Science
Virginia Tech, Blacksburg, VA 24061-0106, USA
e-mail: eales@cs.vt.edu

Laura M. Byrd
Department of Computer Science
Virginia Tech, Blacksburg, VA 24061-0106, USA
e-mail: lmbyrd@vt.edu

Abstract: The Internet has tremendous potential for K-12 education. However, learning how to exploit that potential remains an important problem. In this paper, we use ideas from situated learning and the deschooling movement to address the argument that there has been no significant reform (technology-based or otherwise) of public education for over a century. We present a preliminary educational model focusing attention on the need for engagement with authenticity. We introduce the notion of authentic collaborative learning, and suggest a number of requirements that are desirable in a technological system to support such learning.

Introduction

In this paper, we consider the use of the Internet in K-12 education. Our discussion, at this stage, is primarily theoretical; we are concerned that the educational potential of the Internet is being narrowly defined by technology-driven research based on limited and inappropriate educational models. Our principal interest is in supporting meaningful learning experiences, and we believe the Internet can make a significant contribution toward this goal. We do not want to see the Internet go the way of previous technologies that have promised to positively impact education but have delivered little. It is not sufficient to state, "It will be different this time;" we must demonstrate that it is different and optimize the educational benefits of the difference. We present a preliminary model which focuses attention on the areas in which we believe the Internet can have a positive influence on education.

We are members of a large interdisciplinary project group from Virginia Tech and the Montgomery County Public Schools, supported by a major award from the U. S. National Science Foundation. The Learning in Networked Communities (LiNC) project seeks to exploit the high network bandwidth and availability brought to the County by the Blacksburg Electronic Village (BEV) [Carroll and Rosson 1996] to explore the potential educational uses of a virtual physics laboratory to support collaborative, project-based learning. An important element of the project is to facilitate broad community involvement in education - this is the particular area we want to consider. The views expressed in this paper are our personal views and not those of the project or its participants.

We do not use the terms "learning," "education," and "schooling" interchangeably. We consider their meanings to be quite distinct and these distinctions are important to our discussion.

Is Educational Reform Possible?

"I believe that the motion picture is destined to revolutionize our educational system and that in a few years it will supplant largely, if not entirely, the use of textbooks." - *Thomas Alva Edison*

When a new technology emerges, prominent people tend to rush in to make exaggerated claims about the way that technology will transform education. Such claims can be found for radio, television, computers and now the Internet, in addition to other non-technological educational "breakthroughs." Tyack and Cuban, in their book *Tinkering Toward Utopia* [Tyack and Cuban 1995], present a convincing thesis that for over a century, in the face of a barrage of educational reforms (technological and otherwise), the form and substance of the public education system has remained remarkably stable: "Over long periods of time schools have remained basically similar in their core operation, so much so that these regularities have imprinted themselves on students, educators, and the public as the essential features of a 'real school'." (p. 7)

Tyack and Cuban use historical evidence and case studies to show how it is school that changes reforms rather than reforms that change school. We believe that their argument is an important challenge for all those concerned with the use of the Internet in education. On the whole, technological artifacts play only a small part in education. Educational technology has derived much of its importance from its promise to be able to change education and to provide new opportunities to redress traditional imbalances. This is the promise of the Internet. But it is not sufficient to add another Edison-like statement to the history books, with the proviso that it will be different this time. We need to understand if and why it will be different this time and to focus our research on the issues that are most likely to hold the key to significant educational advances.

Situated Learning

Humans are expert learners. The basic ability to learn has played a significant part in human evolution. We often learn effortlessly without being particularly conscious of what we are doing. Problems develop, however, when we try to control and measure what is being learned. Jean Lave [Lave 1993] reminds us that: "Learning is an integral aspect of activity in and with the world at all times. That learning occurs is not problematic. What is learned is always complexly problematic." (p. 8) A situated approach to learning [Lave and Wenger 1991] focuses on learners and learning. As Seely-Brown and Duguid [Brown & Duguid 1993] point out, "A situated approach contests the assumption that learning is a response to teaching." Learning is embedded in multiple and overlapping social and material situations. These situations are not just a neutral "background" they provide the contextual scaffolding which affords motivation, interpretation, understanding, and so on.

For those that see learning as a fundamentally situated experience, the Tyack and Cuban thesis is strangely reassuring. From a situated learning perspective, the dominant and most enduring experience taking place in schools is "schooling." That is, students learn how to do school: how to pass a test, get a good grade or maybe just survive. In spite of all the fine efforts of teachers to present a systematic and relevant curriculum in stimulating and meaningful ways, it is the game itself that gets into the blood.

One could be encouraged that schools are so successful at teaching schooling. They have a major impact on students' learning - isn't that what they are supposed to do? Unfortunately, evidence suggests that the kind of learning developed and rewarded in schools is very different from the kind of learning that is used and valued outside of school. Lauren Resnick [Resnick 1987] suggests that there are four broad characteristics of mental activity used outside of school that stand in marked contrast to mental activities developed in schools:

1. Individual cognition in school versus shared cognition outside school.
2. Pure mentation in school versus tool manipulation outside school.

3. Symbol manipulation in school versus contextualized reasoning outside school.
4. Generalized learning in school versus situation-specific competencies outside school.

The essential social and material situations of the school have not changed for over a century. Typical standard and universal features of schooling include: age-grading, one teacher per self-contained classroom, full-time attendance, the division of knowledge into subjects, and regular assessment. If we really want to have an influence on educational achievements and, more importantly, on underachievements, we have to do more than just change the curriculum or the medium of delivery. We need to change the fundamental organization of education. We need to break out of the classroom.

Learning Without Schools

If it is schooling that is principally learned in schools, can people be educated without schools? Can we break out of the "school game" and play another game with different rules? Goodman [Goodman 1971] described schooling as a "mass superstition" which nobody opposes and for which nobody proposes alternatives. There have been one or two educational models suggested, however, that are not based on the school. Here we consider the radical educational models proposed by the "deschooling movement" and in particular Ivan Illich [Illich 1973] (see also [Goodman 1971]).

Illich proposed *Learning Webs* as an alternative to schools. He set out to outline the kind of resources required if one considered not what people ought to learn, but instead what kinds of things and people learners might need to be in contact with. He identified four kinds of learning resources: *Things* (educational objects), *Models* (skilled people), *Peers* (other learners), and *Elders* (educators-at-large). Illich also suggested that technology could be harnessed to provide a reference service for these resources.

The great value of Illich's ideas is that he has dared to consider what education might be like without schools. On the other hand, the great weakness of Illich's ideas is that they are difficult to operationalize. It is hard to see how Learning Webs would ever replace the school system. One problem with proposing radical alternatives to schools is that schools have non-educational uses which are very important and have to be considered. Paul Goodman [Goodman 1971] rather cynically describes some of these non-education uses: "In the tender grades, the schools are a baby-sitting service during a period of collapse of the old-style family and during a time of extreme urbanization and urban mobility. In the junior and senior high school grades, they are an arm of the police, providing cops and concentration camps paid for in the budget under the heading of `Board of Education'." (p. 21) Cynicism aside, schools obviously play a central role in our culture.

School systems also represent massive vested interests. They are a substantive part of most of our socio-political and economic structures. It seems ridiculous to propose that we suddenly close the doors to hundreds of thousands of institutions and the people who bring them to life, or to imply that there is any way that we can make a transition to a different way of education without massive upheaval. Although we think Illich's Learning Webs have some value in the context of development of educational resources via the Internet, on this occasion we want to borrow Illich's general notion of deschooling to be carried forward in our argument. For Illich, and for us, deschooling society means far more than just getting rid of the schools; it also means overcoming the schooling mentality throughout the whole of society.

The Educational Potential of the Internet

We believe that if we want the Internet to have a major impact on improving education, the learning involved has to be active and collaborative; but above all, we have to move beyond exclusively school-based conceptions of learning. It will be a significant waste of effort and resources if Internet-based projects only succeed in reifying existing school-based practices or merely "computerizing" limited and simplistic educational models. Indeed, the combination of complex technological models with simple (often tacit) educational models appears to be the best way to negate the educational potential of the Internet (or any other technology). What we need are clear and far-sighted educational models that lay the foundations for a radical and parallel development of learning *and* technology.

There are, of course, limits. We are not suggesting that all learning can be supported by the Internet, or that learning should take place strictly outside the classroom. There are certainly many activities that are best developed in some kind of school setting. The idea we want to emphasize is that the Internet may be a technology uniquely suitable for carrying forward a meaningful integration of learning and community.

An Educational Model

Here we present a preliminary model designed to focus our attention on the "higher" levels of educational activity afforded by the Internet. We consider our model to be cumulative with no clear and absolute boundaries between the different levels - clearly one level merges into the next. We believe such a model to be useful primarily because it draws attention to the third level: the idea of engagement with authenticity. It also cuts across traditional technological boundaries. It is possible to find both simple and advanced examples of technology at each level of the model.

Level 1 - *Engagement with Information.* One of the great values of the Internet, and in particular the World Wide Web, is that it brings the learner face to face (via a fairly standard interface) with an ever expanding universe of digital information. Here the dominant metaphor is the digital library.

Level 2 - *Engagement with Simulation.* Some aspects of the "real world" can never be experienced in a direct sense. Simulation can be of immense educational value in these cases. As collaborative learning can be useful during simulation, it is possible to support collaborative simulation through MUD's (Multi-User Domains) and MOO's (MUD Object-Oriented). Here the dominant metaphor becomes the virtual school or for example, the virtual science lab.

Level 3 - *Engagement with Authenticity.* This is the level which we think is of major significance, particularly in terms of its potential contribution to educational development. It is difficult to think about this area, however, because school has so dominated our educational concepts that it is hard to even find a language in which to discuss the issues. As Illich [Illich 1973] pointed out, "education becomes unworldly and the world becomes non-educational." (p.31)

What we want to facilitate is "virtual access to reality." We consider that the rather ill-defined and somewhat contrived term *authenticity* (meaning authentic activities in authentic contexts) has some value as a general pointer into the issues we need to consider.

Authenticity

Seely Brown, Collins and Duguid [Brown et al. 1989] offer the following definition of authenticity: "The activities of a domain are framed by its culture. Their meaning and purpose are socially constructed through negotiations among present and past members. Activities thus cohere in a way that is, in theory, if not always in practice, accessible to members who move within the social framework. These coherent, meaningful, and purposeful activities are authentic, according to the definition of the term we use here. Authentic activities then, are most simply defined as the ordinary practices of the culture." (p. 34) In an educational context, we use the term *authentic* to refer to activities that, in some way, reach outside of the school community and culture.

Seely Brown, Collins and Duguid's account of authentic activities is fairly representative of the descriptions found in the situated learning literature. An important contribution to the emergence of the situated learning perspective has been the detailed study of learning in traditional or well-established cultures, such as Lave's studies of tailoring in West Africa [Lave and Wenger 1991] or Hutchins' studies of maritime navigation, both traditional [Hutchins 1983] and modern [Hutchins 1993]. Because of these and similar studies, apprenticeship models of learning have received the bulk of the research attention. However, another model of social learning that has received rather less attention is the self-help or mutual aid group, or "collaborative bootstrapping" as we once termed it [Eales & Welsh 1995]. This model is essentially peer-based and, although different members may play different roles and develop different skills, it has very few of the disparities of knowledge and skill associated with the apprenticeship model. In the mutual aid model, it is the motivation to solve a common problem that provides the focus of group activities. One could argue that it is the problem that is authentic rather than some enduring culture. We suggest that this is a far more appropriate social model of learning for use in education and can provide a valuable starting point for the exploration of community-related education via the Internet. We will refer to this kind of learning as *authentic collaborative learning*. This kind of learning would appear to be universal, although it is not often seen in formal education. The author found similar collaborative learning in the face of a common problem amongst administrative computer users in a large Australian university [Eales 1996].

Authentic collaborative learning (via the Internet) will require a certain amount of effort to set up as an educational activity. Educators will have to negotiate access to authentic problems and projects. In particular, interaction with representatives of the wider community is a vital part of the authenticity. Some examples of this interaction, which could be supported by the Internet, include:
with clients - for example, students could negotiate with members of a local community group to create web pages for them.
with advisors - for example, students could seek advice from scientists on the best way to monitor local environmental conditions.
with critics or reviewers - for example, local people could offer their comments on a student created multi-media history of the local community.

Technological Support of Authentic Collaborative Learning

Authentic collaborative learning can and should be supported by technology. At present such activities are usually supported on the Internet by a mixture of e-mail software, web browsers, ftp, word processors, etc. What we require is a simple, robust and integrated tool to support all aspects of authentic collaborative learning. In very simple terms, some of the most important requirements for such a tool are:

- The system should be content-free, although certain kinds of uses may require special methods of capturing and manipulating representations.
- All significant operations of the system should be under the control of the participants.
- There should be support for ongoing, group-based, interactive discourse (usually asynchronous but sometimes synchronous).
- Methods of representation should be appropriate for dealing with the problem but should not require complex skills from the participants (maximal representational value with minimal user effort).
- There should be sufficient media richness to encourage group cohesiveness.
- Adequate privacy and confidence in the security of group boundaries is required.
- Simple and efficient methods of archiving and organizing representations within the group need to be provided.

(For a more detailed analysis, see [Eales 1996])

Conclusions

We have argued that the Internet has a tremendous educational potential but history suggests that this potential will not be realized. Instead of focusing on school-based models of education, the Internet can allow learners to break through the walls of the classroom and engage with authentic activities in authentic contexts. We believe that this approach offers the best opportunity for the Internet to have a significant impact on educational practices and achievements. We have termed this type of activity authentic collaborative learning and have suggested the importance of self-help style group organization focused on authentic problems. Although such learning can be supported by existing Internet-based applications, we have outlined requirements for a system for specifically supporting authentic collaborative learning.

References

[Brown et al. 1989] Brown, J. S., Collins, A., and Duguid, P. (1989). Situated cognition and the culture of learning, *Educational Researcher*, 18(1):32-42.

[Brown and Duguid 1993] Brown, J. S., and Duguid, P. (1993). Stolen Knowledge. *Educational Technology*, 33(3):10-15.

[Carroll and Rosson 1996] Carroll, J. M., and Rosson, M. B. (1996). Developing the Blacksburg Electronic Village. *Communications of the ACM*, 39(12):69-74.

[Eales and Welsh 1995] Eales, R. T. J., and Welsh, J. (1995). Design for Collaborative Learnability. *In the proceedings of CSCL '95*, Indiana University, Bloomington, USA, pp. 99-106.

[Eales 1996] Eales, R. T. J. (1996). *Design for Learnability: Computer supported collaborative learning in the workplace*. Unpublished Ph.D. thesis, University of Queensland, Australia.

[Goodman 1971] Goodman, P. (1971). *Compulsory Miseducation*. Penguin, Harmondsworth, UK.

[Hutchins 1983] Hutchins, E. (1983). Understanding Micronesian Navigation. In Gentner, D., and Stevens, A. L. (eds), *Mental Models*, Erlbaum, Hillsdale, NJ, pp.191-225.

[Hutchins 1993] Hutchins, E. (1993). Learning to Navigate. In Chaiklin, S., and Lave, J. (eds.), *Understanding Practice: Perspectives on activity and context*, Cambridge UP, Cambridge, UK, pp. 35-63.

[Illich 1973] Illich, I. (1973). *Deschooling Society*. Penguin, Harmondsworth, UK.

[Lave and Wenger 1991] Lave, J., and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge UP, Cambridge, UK.

[Lave 1993] Lave, J. (1993). The practice of learning. In Chaiklin, S., and Lave, J. (eds.), *Understanding Practice: Perspectives on activity and context*. Cambridge U. P., Cambridge, UK, pp. 3-32.

[Resnick 1987] Resnick, L. B. (1987). Learning in School and Out. *Educational Researcher*, 16(9):13-20.

[Tyack and Cuban 1995] Tyack, D., and Cuban, L. (1995). *Tinkering Toward Utopia: A century of public school reform*. Harvard University Press, Cambridge, MA.

Acknowledgments

The LiNC project is a joint effort of the Department of Computer Science at Virginia Tech and Montgomery County Public Schools. It is funded by grant REC-9554206 from the U. S. National Science Foundation, under the Networking Infrastructure for Education program. Additional support is provided by Apple Computers. We thank our fellow project participants, especially Kathy Bunn, Mark Freeman, Suzan Mauney, and Fred Rencsok, all teachers in Montgomery County Public Schools, for their fellowship and support.

Medtec: A Web-Based Intelligent Tutor for Basic Anatomy

Christopher R. Eliot
Daniel E. Neiman
Michelle M. LaMar
Center for Knowledge Communications
Department of Computer Science
University of Massachusetts
Amherst, MA 01003
{eliot, dann, lamar}@cs.umass.edu

Abstract: We present our experiences in developing Medtec, a web-based intelligent tutor in the domain of basic anatomy and physiology developed as a prototype for training Air Force Reserve medical personnel. We discuss some of the challenges encountered in performing intelligent tutoring over the web, including generating topic-based hypertext from a previously existing course, creating flexible interactive mechanisms for drilling and testing student knowledge, and acquiring sufficient feedback to develop useful student models.

1 Introduction

The research mission of the Center for Knowledge Communications at the University of Massachusetts is to advance the theory, development and deployment of intelligent tutoring system technology. An intelligent tutoring system actively monitors a student's progress, maintains student models, and reasons about the best way to present material. These tutors often include substantial multimedia components consisting of graphics, animation and digitized sound. The resulting systems are both memory and compute intensive, demanding resources that exceed those generally available to our target audience – mostly educational facilities. Since we must target computational resources already available to the students, cross-platform delivery is essential. Even assuming that appropriate hardware is available, installing and maintaining prototype research software in an educational environment at any distance from our research facilities is burdensome. Collecting data from student activities distributed across multiple sites is also difficult.

In short, intelligent tutoring systems impose a number of constraints upon the delivery systems that are admirably met by a web-based technology. The web provides a universal cross platform interface while the requirement that the platform support a Java-enabled web browser can be easily described and understood. Intensive computation and reasoning activities can be performed on the server in our research lab which we can maintain more easily than client machines in the student's institution.

The main drawbacks of a web-based tutor that we have found are reduced interactivity and the limited view of the student's activity available to the server through the limited vocabulary of http requests. We present our experiences in developing a web-based intelligent tutor and discuss how we address some of these obstacles in the context of Medtec, a tutor in the domain of basic anatomy and physiology developed as a prototype for training Air Force Reserve medical personnel.

2 The Medtec Tutor

The Medtec Tutor presents a basic course in physiology and anatomy based upon the complete pre-existing text of the Air Force Reserve Anatomy and Physiology training manual [ES]. We have enhanced the text with many active study aids implemented using various Web-technologies including Java and hyperlinks. Adding color and indexing to the material increases its appeal significantly. Computer graded self-tests and anatomy drills give students instant feedback concerning their degree of understanding. Semantic links among learning activities allow the tutor to provide students with guidance and support better management of their study time.

Structural and functional knowledge are presented to the student using any of several methods – textual and graphics presentations, as well as animation. Pedagogical scaffolding includes tracking of student progress, and presentation of appropriate support and drill material. The student is guided, according to student and domain models, to the appropriate subjects based on mastery and logical structure of the domain. The student model is partitioned into the general topic areas and records material encountered, tutoring methods applied for each topic, mastery tests applied and degree of proficiency.

2.1 The CL-HTTP Server

The server for the Medtec Tutor is the Common Lisp HTTP server developed by John C. Mallery at MIT [Mal94, Mal]. We chose this server for several reasons, the primary one being that we could then implement our web-page generation software in the same language as our reasoning and student modelling software. We also take advantage of CL-HTTP's facilities for handling forms without CGIs and for dynamically generating html pages and exporting URLs.

CL-HTTP, by virtue of being integrated with Common Lisp, supports richly structured knowledge representations extremely well, without the overhead implied by external database representations. Perhaps more importantly, the fundamental paradym of CL-HTTP is more suited for tutoring applications than traditional web-server architectures. Traditional web-servers are oriented towards static pages with extended features providing some dynamic content. CL-HTTP, on the other hand, encourages dynamic computation of page content in preference to static pages.

Creating all pages dynamically provides more ability to reason about content over form and enhances stylistic uniformity. Rather than cutting and pasting standard design elements into each page, we generate standard design elements programmatically. This supports uniformity of design without limiting evolution. If some standard design element is revised or refined, then only a single function must be changed and all pages using that design element are instantly revised; a traditional site with many pages sharing copies of that design element would require edits to many pages.

2.2 Partitioning the Text for Presentation and Tutoring

The body of the course material is based on a pre-existing text on basic anatomy and physiology. In converting the text to an HTML format for tutoring, we had multiple, not necessarily compatible, goals. For student modeling purposes, we needed to be able to associate text with topics. This allows us to create a model of what the student knows, based on the text that has been visited, and to present the appropriate text when some form of interaction has indicated that the student does not have full command of that topic. Our goal is to make the system highly adaptive so that students are presented with material selected to be most appropriate to support their learning.

However, the presentation must also provide students with a comprehensible sense of context. A continuous stream of tutoring material pushed onto the computer screen would quickly confuse any student. Consequently, we chose to retain the book metaphor. Students may always select a chapter, section, or subsection so that the student has a familiar orientation from which to study the material. We adapt to the the student's performance by sharing control over the student's location within a fixed interface metaphor. Review sections at the end of each chapter provide adaptive content spontaneously generated based upon the student model.

2.2.1 Indexing Topics

Because we do not have substantial natural language capabilities in our system, we indexed the textbook material as topics at the paragraph level. Each topic includes a learning (or presentation) method, previous, next, parent and children links, and variables for student modeling. Precondition and postcondition links are intended to support sophisticated reasoning but have not currently been implemented because of the knowledge-engineering effort required to aquire precondition knowledge for every topic in the textbook. Simple navigation through the textbook metaphor is supported by the previous, next, parent and children links. Lexical links between words and topics are also implemented. The server includes both a stored dictionary and dynamic indexing of every word used in the text. Search facilities allow students to move directly between alphabetic word indexes and topics.

3 Interactions with the Medtec Tutor

Our approach to making the Medtec tutor interactive and engaging was multi-fold. We added basic form-based test methods, Java-based training tools, and adaptive review pages that generate tests based on student's past performance.

The most basic level of interaction was to implement HTTP forms for handling basic knowledge queries such as multiple choice questions and column matching. (Short answers are problematic because they require some form of natural language parsing ability.)

3.1 Active Diagrams

The original text contained a large number of figures consisting of labelled diagrams, for example, a schematic of the heart showing the valves and chambers. We replaced these figures with active diagrams – graphical presentations with multiple modes of interaction. Students can use these diagrams in several modes: to study the whole diagram, to focus on individual terms, and to drill themselves on the diagram content. In the basic learning mode, the diagrams are presented in a manner similar to a text figure, with labels connecting terms to their corresponding representations. We have experimented with several small HCI innovations in this mode – the color or shading of the text may indicate the degree to which the tutor is confident that the student knows the term based on the result of recent drills and, when the label is mouse-buttoned, the corresponding item might be highlighted or a definition might be presented.

This standard layout, however, can overwhelm a student to whom all the information is new. We have therefore provided a flash card mode that focuses the student's attention on one feature at a time. In this mode one feature of the diagram is highlighted (e.g. the aorta), and its label alone is displayed for a few seconds, after which the applet will automatically switch to a new feature, focusing on each in turn.

The quiz or drill mode allows a student to test their own retention of the terms as well as providing feedback to the server on the student's progress. The student is quizzed both on feature recognition and feature recall. For recognition we highlight a feature of the diagram and ask the student to identify its label from a multiple choice list. To test recall, we ask the student to click on the location of a feature in the diagram given its label.

3.1.1 Implementation of Active Diagrams

Our initial implementation of active diagrams was built using a Shockwaved Director movie.[1] This approach proved unsatisfactory because there was no non-trivial way to return the results of the student interaction to the server. Active diagrams are now implemented using a Java applet connected to the HTTP server via a two-way socket. The applet displays data transmitted by the server in the appropriate mode (after initially loading a situation-specific set of gifs and data points).

Because the server is responsible for performing all the reasoning for generating tests and for recording results, the resulting Java applet is small (reducing download time) and general (the same applet can be used for virtually all active diagrams). Decoupling the reasoning component from the actual quizzing and presentation component gives us greater flexibility in modifying and experimenting with different tutoring strategies.

3.2 Feedback based on Student Modeling

In the Medtec Tutor, a Markov algorithm is used to focus student's attention. Students may review each topic as often as desired and the system records a timestamp for each student interaction with each topic. The system separately maintains an evaluation of the student's knowledge of each topic. Since this domain primarily consists of memorization, this evaluation is a simple numeric score; more complex domains will require more complex representations. Using a simple markov process the system computes a reinforcement interval for each topic, based upon the student's level of understanding. When students do not understand a topic at all, the system's goal is to provide as much practice as possible on the topic. When the student's understanding increases, the system's goal transitions to ensuring retention; topics that are considered to be well-understood are still presented, but at increasing intervals.

---

[1] Shockwave and Director are trademarks of Macromedia, Inc.

The target interval for lessons on a topic is interpreted as a system recommendation, not an overriding priority. This recommendation is integrated with other considerations in flexible ways. The student may always review any topic, regardless of the system's belief about which topics the student will best learn from; the system never overrides a direct student goal. However, chapter review and self-test sections are generated dynamically on the basis of system controlled student models. In addition, if the student asks for guidance, the student model is always available as a basis for computing directions. The result is a mixed-initiative pedagogy for curriculum control.

4 Limitations of Web-Based Tutors

The primary difficulty in implementing web-based tutors is the lack of flexibility provided by the HTTP vocabulary and lack of interactivity and programmability in the browsers themselves. Although browsers are very good at presenting text, the stateless nature of the HTTP protocol is designed to prevent the very kind of feedback regarding browser activity that we need to track student progress.

The most difficult aspect of implementing a tutor using web browsers is monitoring the student's progress. Although we can present text to the student, there is no way of directly confirming that the student has read it. Browsers do not report page level activities back to the server, thus we can not even confirm that the student has scrolled down a page encountering each section. Certainly, we could attempt to structure the text in such a way that each page corresponds to a single unit of knowledge, however this would require decomposing the text to the point where each page consisted only of a single sentence or, at best, a paragraph.

Instead, we must determine the student's progress indirectly using the aforementioned interaction mechanisms and connecting them with the appropriate subjects in the text and making other indirect inferences based, for example, on the amount of time spent on each page and the number of times that a particular page has been visited.

Implementing student-side tutoring aids is also made difficult due to the lack of information provided by the browser. It is difficult, for example, to construct a notebook facility in which the student can refer to, or annotate, specific sections of text at a level more specific than that provided by the page bookmarking capability of most browsers.

A complete solution to the issues we have raised here would probably involve the implementation of a Java-based text browser that would allow us to directly access the text and material being viewed. However, this approach involves a degree of implementation that we greet with little enthusiasm. Instead, while we wait for the standards to catch up with our needs, we have improvised mechanisms for increasing the accuracy of our views of students's activities. An invisible Java applet,for example, records when a student departs from a page (thus preventing the case in which a student jumps from a tutor page to a page outside the domain, corrupting our timing results). Our division of the text into discrete topics allows us to generate and jump to semantically meaningful divisions in the text.

5 Conclusions

By integrating Java applets with a web server implemented using CL-HTTP we have achieved substantial student interactivity in our web-based tutor. We believe that this technology makes the web a viable and desirable mechanism for delivery of high quality educational technology. Although programming in this environment requires some restructuring compared with traditional methods for implementating intelligent tutoring systems, all of the fundamental concepts of student modeling and knowledge-based domain representation can be exploited in the web-based context.

The structure of the web is adequate to support high quality tutoring systems, but performance is still a limiting factor. Experiments with Java applets reveal that runtime performance is limited. Better optimization of the Java virtual machine and fully compiled Java implementations should soon be available to address this problem. Network response time and transmission speed must be considered, but for tutoring these factors may be less important than for other applications. We intend for students to study our web pages, not to browse them and hence we can accept some delays while moving from one page to another.

Performance of the server is a third consideration. We do not have experience with large-scale usage of our system yet, so we have no definitive data about server performance. The experience we do have suggests that server performance with CL-HTTP is good. If server performance becomes an issue, we have many options for hardware improvements. Since problems with server performance directly relate to the size of the user-base, it is reasonable to rely on hardware improvements for maintaining adequate performance because the cost of additional hardware can be distributed among many affected users.

Despite the limitations of web-based tutors involving the limited view of students' activities, we have found the technology to be of great benefit in developing, fielding, and maintaining the Medtec tutor and we believe that our experiences will generalize to many of the other tutoring efforts in which we are currently engaged.

Acknowledgements

This work was funded under NCC OSF #N66001-96-C-8604.

References

[ES] MSgt Beverly A. Ellard and Nancy F. Smith. *CDC90250A, Medical Service Specialist, Volume 2, Anatomy and Physiology.* Extension Course Institute (AU), Maxwell Air Force Base (Gunter Annex), Alabama, 36118–5643.

[Mal] John C. Mallery. Common lisp hypermedia server (cl-http). Web reference: <http://www.ai.mit.edu/projects/iiip/doc/cl-http/home-page.html>.

[Mal94] John C. Mallery. A common lisp hypermedia server. In *Proceedings of The First International Conference on The World-Wide Web*, May 1994.

It's life Jim, but not as we know it!

Henrik Fagrell
Department of Informatics, Göteborg University, Sweden
E-mail: fagrell@adb.gu.se

Carsten Sørensen
Department of Computer Science & Applied Mathematics, Aston University, UK, and
Department of Informatics, Göteborg University, Sweden
E-mail: c.sorensen@aston.ac.uk and carsten@adb.gu.se

Abstract:

The World Wide Web (the Web) is currently the main driving force behind the rapid diffusion of Internet technology in professional and private contexts. This rapid development leads to that more and more people living a larger and larger proportion of their lives in Cyberspace. Mea suring and monitoring our natural and artificial surroundings are crucial human activities in order for us to both understand and intervene. Very little research has, however, investigated how we can survey and analyze the Web. This paper explores how we can collect data about the Web in order to be able to describe and compare measurable aspects such as; topology, language, layout, size, and density. We ask the following question: *How can we apply semiautomatic quantitative measurement instruments to better understand changes to the contents of the Web?* A host of Web measures are suggested and results from a Web survey conducted by software robots are presented and discussed. It is concluded that this approach can inform future design of a number of specialized Web services such as search engines and advertisement.

1 Introduction

The recent explosive diffusion of the Internet and the World Wide Web [Berners-Lee et al. 1994] has led to an increasing amount of both professional and everyday life being carried out in Cyberspace. Many of us spend much time every day in this new 'world' of bits [Mitchell 1995], but what do we really know about it? In the physical world we have been monitoring an abundance of both cultural and natural attributes through hundreds of years. Monitoring and measuring are fundamental activities for understanding the world we inhabit and shape. We must, therefore, also develop ways of measuring and monitoring the World Wide Web.

The Web contains a vast amount of hypertext. This body of information can be navigated using a bottom-up search through the floodgates of a Web search engine or top-down through tiny portholes of index pages providing taxonomies for Web contents. We do, however, know very little about this new cultural phenomenon. In the words of Dr. McCoy to Captain James T. Kirk of the Star ship Enterprise, faced with a brown blob of unknown origin, "It is life Jim, but not as we know it!"

Several research efforts have attempted to alleviate this. Some have studied how we can represent, visualize and analyze quantitative attributes of the Internet, such as traffic patterns and size [Rickard 1995; Blundon 1996]. Others have investigated how we can study the *use* of the Web, i.e., relationships between people using the Web and the contents of it [Hoffman et al. 1996], and a several efforts have explored how to map Web hypertext structures, how to perform analyses of link topologies between Web-sites [Drew et al. 1995; Girardin 1996; Mukherjea and Foley 1995], and how to apply statistical analyses of the volume and density of the Web [Bray 1996]. Yet next to no research has explored how to measure and analyze the contents of the Web. If the physical world is viewed as a world of atoms, and the World Wide Web is seen as a 'world' of bits [Mitchell 1995; Negroponte 1995], then the question asked in this paper is: *How can we apply semiautomatic quantitative measurement instruments to better understand changes to the contents of the Web?*

The paper suggests five different types of Web measures: volume, density, vocabulary, structural and relative. The feasibility of applying semiautomatic Web robots to collect data about changes to the Web contents is investigated. In the scope of this paper we are interested in what we can learn about the Web from studying, for example, the distribution of Java Applets on Web sites, the average number of errors per Web page, or differences in vocabulary between two Web sites. We are primarily interested in assessing the feasibility of studying content changes to Web pages. We are at this stage not interested in how to study what people think, feel and say when they surf the Web. The perspective adopted in this paper does not accommodate current trends focusing on "push" technologies based on basic technologies such as ActiveX, Java and Castanet. Here, channels transmitting information replace the text and document metaphor of the Web. Studying the two phenomena will require different approaches. However, as noted by [Kelly and Wolf 1997], the emergence of such technologies will not replace the Web, merely supplement it.

In order to explore the question, we have conducted a survey of 82 Web sites within the ".se" domain, i.e. on Swedish Web sites. A software robot, named "*Ethel the Aardvark*", were designed, implemented and tested, and used in the survey. Configuring the robots with a list of sites to be visited was a manual task, while data collection and data aggregation were automatic and semi-automatic. The research approach can be described in terms of the following activities: (1) specification of Web measures to be calculated; (2) design and construction of software robots; (3) small-scale tests of robots; (4) selection of robot for survey; (5) web-site selection; (6) data collection; (7) data aggregation; (8) data analysis; and (9) documentation of results. The survey illustrated the feasibility of conducting surveys of the Web, and we illustrate this by results from analyzes of Swedish newspaper sites. It is concluded that this approach can inform future design of a number of specialized Web services such as search engines, Web advertisement, and a Web Dow Jones Index.

In the following section we survey related research and outline the problem setting. Section 3 suggests a host of quantitative measures for surveying the Web. Section 4 presents the instruments and procedures for collecting and analyzing data in the survey. Section 5 presents the results from survey applying Web measures from each of the categories suggested in Section 3. Section 6 concludes the paper and discusses practical applications of this approach.

2 Surveying a Brand New World

How can we survey the Web consisting of a dense weave of texts, pictures, interactive components, CGI scripts etc? Within a relative short time span a number of different approaches have been suggested for studying the Internet in general and the Web in particular. This section presents our approach measuring the Web and relates it to similar approaches. It is beyond the scope of this paper to list all related research. [Dodge 1997], however, presents the most comprehensive list of references we have found, and his index has proved valuable when gaining an overview. As in Dodge's list, some of the references in this section are URL's, simply because this is where the information is available. We have, however, sought to replace as many as possible of these with references to refereed material.

Viewing the Web as a 'world' of bits naturally raises the issue of space. In geometry, space is defined by two concepts: topology and metric. If we use the geometrical definition of space as a metaphor the Web's topology can loosely be describes as a graph with nodes and directed links. The nodes can be represented by retrievable documents, that is, files containing texts, images, links, and several other types of information. A plain distance metric does not capture the phenomenon accurately. Increased physical distance between the computers connected in the network does not necessarily lead to higher transaction costs. The metrical aspects can, however, be based on other variables than distance. Other researchers using the geometric metaphor considers the Web's metric to be calculations on how to traverse the graph formed by the link structure [Drew et al. 1995; Girardin 1996; Mukherjea and Foley 1995].

Because of the magnitude of the task of surveying the content of a large proportion of the Web, it is extremely important to stratify the sampling, i.e. select a target population. The survey, therefore focuses on collecting and analyzing data from Swedish Web-sites within relatively few sectors, e.g., newspapers, companies registered on the stock-exchange and government agencies. This is in no way different than attempts to survey the physical world. Geographers, sociologists, economists and statisticians are also forced to stratify their areas of inquiry. The annual Swedish statistics report [SCB 1996], for example, only contains a tiny fraction of attributes measured, which in turn only represents an infinite fraction of the attributes measurable. It is also important to stress that this paper only attempts to investigate how to study the Web from the publicly accessible side. We intend study what is inside the Web, which is different from a number of other approaches which analyze aspects of Web sites which is not publicly available, such as, activity logs and restricted access-areas.

A major part of the research on both the Internet and the Web suggests approaches mapping Web-user demographics and behavior. [Pitkow and Kehoe 1996] present a series of comprehensive demographic surveys of Web-use patterns, e.g., average age of users and gender, conducted by researchers at Georgia Tech and others [Hoffman et al. 1996] presents a study of the use of the Web. The authors investigate Web-use patterns in the area of electronic commerce by analyzing data from the CommerceNet/Nielsen Internet Demographic Survey [COMNET 1996] sampling questionnaire data on Web-use from users. A more relevant strand of research is concerned with studying how to map and visualize both the Internet and the World Wide Web in order to provide support for navigation, and addresses issues such as: maps of the Internet, Internet repositories and indices, statistics of Internet traffic and size, and visualization of Web spaces [Dodge 1997]. The following provides a few examples of this type of research. [Girardin 1996] and [Drew and Hendley 1995] are mostly interested in visualizing hyper-link structures. This is just an example of a number of research efforts that attempts to visualize information and not survey. [Barry and Batty 1994] analyze the diffusion of the Internet in order to predict future growth. Dodge applies a spatial metaphor to analyze the Web using Geographical Information System (GIS) technology.

Bray suggests collecting data and performing statistical analyses on volume and density measures of the Web [Bray 1996]. The project, furthermore, looks at the relative link topology between Web sites. Bray applies software robots for automatically collecting data. This approach has a number of similarities to the approach we suggest, but there are also major differences. Bray's survey of the Web is based on the Open Text Index, November 1995, covering 1,5 million pages. The parameters analyzed are, however, quite few and they are mainly volume and density measures, e.g., distribution of page sizes, number of embedded images, and types of file extensions. These are combined with structural measures such as a ranking of sites most often referred to, and other inter-site linking measures. The inter-linking measures are applied to illustrate proximity of sites through a spatial mapping.

Bray's approach and the one adopted in this paper both apply the Web site and page as the two basic sample units. It could be argued that surveying the Web based on site names defining the granularity is biased. By putting the site in center we focus on

BEST COPY AVAILABLE

institutionalized entities on the Web. One way of taking this into consideration is to calculate "links-to-site" sets. Bray calculates rankings of most popular site referenced to in the pages. While Bray focus on few and relatively simple measures for a large sample, we have chosen to measure more parameters, and to focus on analyzing the contents of the pages deeper. Since the Web today is a rapidly growing body of hypertext we have found it highly relevant to augment the existing body of research with an investigation of the feasibility of contents-based analysis. The aim is, furthermore, to suggest an initial classification of Web measures. [Table 1] outlines the perspective adopted, and the following section presents five different types of Web measures.

What is surveyed?	The publicly accessible parts of the Web
What is the unit of sampling?	Strata of Web sites
Which collection method is used?	Web robots
What is sampled?	The hypertext contents of Web pages

Table 1: The basic approach explored in this paper.

3 Web Measures

A hypertext contains two fundamentally different types of data: the content and the tags which is meta-data describing the layout and linking structure between the text, graphics, audio and interactive components. Analyses of hypertexts, therefore, concern both aforementioned types of data, and in this paper we suggest the application of five different types of quantitative measures drawn basic measurement and linguistics literature [Voionmaa 1993; Tesitelová 1992] [see Table 2].

Measures	Description	Examples
Volume	Count absolute numbers of hypertext atoms (e.g. the tags and the text). This constitutes all raw data collected from which the remaining measures are calculated.	Number of separated word (tokens), different words (types) when attempting to access the page.
Density	Density calculations on the volume measures.	Number of errors pr. page. The standard deviation for tokens pr. page.
Vocabulary	Identifies the richness of the used vocabulary.	Vocabulary measures such as Guiraud or theoretical vocabulary can be used
Structural	Attempts to measure the site hierarchy, depth and width of the link tree.	The average number of clicks on internal links needed to get from the start page to any other page.
Relative	Compare different data sets.	Lexical equality measure identifies whether two texts deal with the same topic, or have a similar content.

Table 2: A quantitative survey of the Web can be conducted applying the following five different categories of quantitative measures.

Volume measures count total numbers of constituents in the hypertext: bytes, pages, link errors, tokens, types, headings, interactivity, internal and external links. The number of bytes and pages provide measures of the size of a site. The number of link errors reflects how well administrated it is. The total number of tokens (separated words) and types (different words) provide contents-based volume metrics for a site. Interactivity is measured by counting forms, CGI-script and Java-applets. Measuring headings, external links (to other sites) and internal links (within the site) provides quantitative measures for 'page-layouts'.

Density measures relate volume measures to each other, making it possible to express more general site properties. Examples of density measures are: Bytes pr. page, average number of tokens pr. link error, and number of external links per page.

Vocabulary measures analyze site text vocabulary applying the linguistic measures: Guiraud and theoretical vocabulary [Voionmaa 1993; Tesitelová 1992]. Guiraud is a measure reflecting vocabulary richness. It is calculated by dividing the number of types by the square rooted number of tokens. This measure does not incorporate the size of the corpus, and subsequently fails on both extremely small and large texts. Because of the large variations in the size of Web sites we have used theoretical vocabulary as a complement to Guiraud. Theoretical vocabulary is not sensitive to the corpus size, but because it is computed based on a frequency list of types, it is computationally more complex than Guiraud.

Theoretical vocabulary reflects the expected number of types if the tokens are reduced. The measure is calculated as follows: Suppose that a text containing N number of tokens should be reduces to M number of tokens. Let V be the number of word types. The possibility that all occurrences of a word type gets lost in a reduction is $(M/N)^i$. If T_N is the original number of types the theoretical vocabulary will be (T_M) [see Figure 1].

$$T_M = T_N - \sum_V V_i \left(\frac{M}{N} \right)^i$$

Figure 1: The theoretical vocabulary formula.

We reduced the number of types (M) to ten thousand. Both Guiraud and theoretical vocabulary values increases when the vocabulary gets richer.

Structural measures provide quantitative measures representing the spatial property distance. At this stage, we only suggest mean distance as a structural measure. It reflects whether the site link-structure is deep or flat by giving the average number of the smallest amount of clicks on links needed to get from the start page to any other page.

Relative measures compile various differences between sites. We use lexical equality as a relative measure to detect if two sites use the same type of language. Lexical equality is calculated from the words used in the texts without considering where in the text the words are, and this is accomplished by using frequency lists. Lexical equality can be calculated in two ways: token-based or types-based. The types-based method does not consider the frequency of the word, e.g., two texts containing the same words are despite size considered equal. With the token-based method the frequencies of words are taken into consideration, but the problem with it is that context carrying words often have low frequencies. Context carrying world are often nouns or verbs and explains more about the texts than highly frequent words such as "*and*", "*or*" and "*I*" do. Lexical equality is expressed as a percentage. When lexical equality has been calculated for every combination of sites, the values are put in a matrix that gives a lexical distance map. The values in the matrix are then visualized using clustering [Jain and Dubes 1988]. Clustering on lexical equality on conventional newspaper articles has earlier been done by Hagman and Ljungberg providing interesting results [Hagman and Ljungberg 1995]. Clustering gives us the possibility to find patterns in how the sites are lexically related to each other. Olsen has used clustering to get an overview of a hypertext document collection. Here, the clustering was based on the keywords of each document [Olsen et al. 1993].

BEST COPY AVAILABLE

The basic survey features of some of the more advanced Web search engines, e.g. AltaVista and Lycos, collect similar information in a similar way to this project but with completely different purposes, namely that of building key-word indexes in order to facilitate information retrieval. The search engine collect lexical data in order to index words and sentences but do not compile data on the vocabulary, the site structure or relations between sites.

4 Survey Setting

All of the sites surveyed were found in the Swedish University Network (SUNET) link collection at URL: http://www.sunet.se/sweden/main-sv.html. We do not have resources to investigate the entire Web, and we, furthermore, intend to use our contextual knowledge about the selected sub-strata during the analysis. Although all sites are Swedish with server address in the ".se"-domain, this is no guarantee for the site physically being located in Sweden. For example, the server *www.ericsson.se* seems to be physically located in the Netherlands. The survey was conducted on 82 Web sites and the following activities describe the survey: (1) specification of Web measures to be calculated; (2) design and construction of software robots; (3) small-scale tests of robots; (4) selection of robot for survey; (5) web-site selection; (6) data collection; (7) data aggregation; (8) data analysis; and (9) documentation of results.

To collect data from the Web we used web robots, which also are referred to as 'Web Wanderers', 'Web Crawlers', or 'Spiders'. These names are, however, misleading as they give the impression that the software itself moves between sites like a virus. This not the case, a web robot simply visits sites by requesting documents from them. Web robots are also used by the search engines (e.g. AltaVista, Lycos) to collect data for indexing.

A (web) robot is a program that automatically traverses the Web's hypertext structure by retrieving a document, and recursively retrieving all documents referenced. The term 'recursively' does not limit the definition to any specific traversal algorithm. The robot can apply some heuristic algorithm to the selection and order of documents to visit, it is still just a robot. A web browser is not in itself a robot since it is operated by a human user and does not automatically retrieve referenced documents. If the robots do not contain rules stipulating when to stop, it might attempt to retrieve all the public pages on the Web. This could for example happen when the robot has reached a certain depth in the link structure, or when a predefined number of documents have been retrieved. The criterion applied in our experiments are defined by all the public pages within a given site or domain.

A robot has to behave according to ethical rules [Eichmann 1994; Koster 1995], such as avoiding to squire resources from human users by retrieving pages at high speed. It must also identify itself to the web server so that the webmaster can contact the owner of the robot if problems occur. An example of such a problem might be when the robot is getting stuck in a 'black hole' which is a page with a script designed to generate a new page when accessed. This detains the robot until its owner shuts it down, possible after it has caused nasty network delays or finished of a disk or two.

Three different robot prototypes were constructed. The first one was an extension of the maintenance robot MOMSpider [Fielding 1994], implemented in perl and used for validating links and generating statistics. Due to performance problems with perl, a second robot was developed in C++. During the development of the second robot we came across ht://Dig (available at URL http://htdig.sdsu.edu/) implemented in C++. It is constructed to index local networks, such as Intranets, but with some adjustments it served our purpose perfectly. We named our tailored version of ht://Dig "*Ethel the Aardvark*" [Monty Python 1980] and all the data documented in this paper is collected by Ethel.

BEST COPY AVAILABLE

When the design, construction, test of the robot was finished we started to collect data from the selected Web sites out of 6 different sectors [see Table 3].

Sector	Count	%
A-List, i.e., companies registered on the Swedish stock-exchange	24	29,3
Municipalities	11	13,4
Newspapers	8	9,76
Political parties and interest groups	13	15,9
Government agencies	19	23,2
TV- and radio stations	7	8,54

Table 3: Frequencies and percentages of the sites analyzed.

[Table 4] shows key sampling data on the total amount of hypertext sampled. It also provides information on size, download time, number of tokens and types, and the calculation time for the frequency lists for both the largest and the smallest Web site.

All sites	310 mega-bytes uncompressed hypertext. 21 mega-byte URL-lists
Largest site and frequency list	Ericsson, Oct 14. 17,500 kilo bytes hypertext downloaded in 10 hours. 2000 kilo bytes frequency list. 2,489,999 tokens and 163,636 types calculated in 3500 seconds (with a optimized C-program)
Smallest site and frequency list	Dagens Industri, Oct 14. 178 kilo bytes hypertext downloaded in 12 minutes. 17 kilo bytes frequency list. 11,014 tokens and 1,948 types calculated in 7 seconds (with a optimized C-program)

Table 4: Key sampling data on the total amount of hypertext sampled: the largest site and frequency list, as well as the smallest site and frequency list.

In order to reach a sufficient depth in the analysis, we chose to focus on one of the sectors surveyed— newspaper Web-sites. In general, they change more frequently compared to other categories, and since we intend to monitor changes in the data sets, they are the most appropriate. As an example, there were no changes during the sample period on any of the A-list companies' Web-sites. The newspapers sites were collected at five different occasions in 1996, namely, September 23rd, September 30th, October 14th, October 29th and November 4th. The newspapers are all in the daily press, and are listed in [Table 5].

Newspapers	**Description**	**Web-site started**
Aftonbladet	National evening paper	August 94
Arbetet Nyheterna	Regional morning paper	March 96
Dagens Industri	National business daily	June 95
Göteborgs Posten	Regional morning paper	August 95
Hallandsposten	Regional morning paper	September 95
Nerikes Allehanda	Regional morning paper	May 95
Sydsvenska Dagbladet	Regional morning paper	August 95
Svenska Dagbladet	National morning paper	June 95

Table 5: The Swedish newspaper Web sites analyzed, with a indication of the month when the Web service had been launched.

In order to obtain more information about the newspapers, the organization Tidningsstatistik AB /Reklamstatistik AB (TSRS) was contacted. TSRS is a member of the International Federation of Audit Bureaux of Circulations, and it examines and revises newspapers (URL: http://www.tsrs.se/).

The data aggregation was conducted with a variety of small programs implemented in several different languages, e.g., C, perl and awk. Part of the data aggregation process was automated by perl scripts 'gluing' the various programs together. Standard statistical packages (DataDesk and Microsoft Excel) were used for calculations and hypotheses testing. We have also used data clustering [Jain and Dubes 1988] to visualize relative results in order to establish patterns in the data material.

5 Results

The following presents two examples of analyses of data from the survey, applying techniques described in Section 4, and the measures described in Section 3: (1) comparing detailed data from two different newspaper-sites, and (2) a cluster analysis of lexical equality of all newspaper sites.

Analyzing Individual Newspaper Sites

Göteborgs Posten (GP) is geographically located in Göteborg and operates in the western part of Sweden. GP is the second largest morning newspaper in Scandinavia with an average circulation of 273,600 on weekdays and 306,700 on Sundays. [Table 6] shows the five data-samples from Göteborgs Posten's site. Sydsvenska Dagbladet (SD) is also a regional morning paper [see Table 7]. Both GP and SD initiated a web sited in August, 1995.

Measures	I: 23/9	II: 30/9	III: 14/10	IV:29/10	V: 4/11	Average	Std.dev
Bytes	5113090	5152952	5209807	5629806	5980213	5417173	376665
Pages	912	914	929	902	943	920	16
Tokens	502809	509056	512421	569376	606736	641609	45874
Types	58210	58661	59171	59102	61812	59917	1407
Link error	5	4	3	85	91	38	46
Links	9943	10083	10229	9719	10351	10065	247
Headings	1574	1541	1543	1494	1605	1551	41
Bytes/pg	5606.46	5637.80	5607.97	6241.47	6341.69	5887.08	371.16
Tokens/pg	1107.96	556.95	551.58	631.24	643.41	698.23	282.83
Types/pg	66.71	64.18	63.69	65.52	65.55	65.13	1.20
Link err./pg	0.0055	0.0044	0.0032	0.0942	0.0965	0.0408	0.0498
Links/pg	10.90	11.03	11.01	10.77	10.98	10.94	0.011
Headings/pg	1.73	1.69	1.66	1.66	1.70	1.69	0.03
Largest page	35449	43080	40376	44924	51516	43069	5918

Guiraud	58.0	58.1	58.5	55.4	56.1	54.2	1.4
Theor. vocab.	2455	2454	2458	2409	2410	2437	25
Mean Dist.	2.3	2.3	2.3	6.6	6.7	4.0	2.4

Table 6: All key data from the five samples of the Göteborgs Posten web-site.

The standard deviation of the size of Sydsvenska Dagbladet's site is bigger then many of the other sites, such as Hallandsposten with a maximum of 601,781 bytes, Dagens industri with 175,748 bytes, and Arbetet 621,353 bytes. The theoretical vocabulary gives a quantitative measure of the diversity of a text. The average for all sites is 1819 and the top score is 2617 (The Royal Library). This makes the 2437 average for Göteborgs Posten, and 2405 for Sydsvenska Dagbladet quite high. The diversity of the language seem to improve some over time, whereas the average page size seems to decrease. Apart from changes to the vocabulary, no radical changes seem to have taken place on the Sydsvenska Dagbladet's site.

Measures	I: 23/9	II: 30/9	III: 14/10	IV: 29/10	V: 4/11	Average	Std.dev
Bytes	12887420	13130218	13854795	14316155	14437058	13725129	694378
Pages	1327	1353	1438	1520	1543	1436	96.5
Tokens	1550271	1574274	1642583	1700818	1713143	1636218	73066
Types	124015	125301	128940	131420	132127	128361	3710
Link error	21	18	17	22	21	20	2.17
Links	11052	11271	13095	13852	14042	12662	1417
Headings	1855	1876	1820	1930	1950	1886	53.5
Bytes/pg	9711.70	9704.52	9634.77	9418.52	9356.49	9565.20	166.42
Tokens/pg	1168.25	1163.54	1142.27	1118.96	1110.27	1140.66	25.89
Types/pg	93.46	92.61	89.67	86.46	85.63	89.56	3.52
Link err./pg	0.0158	0.0133	0.0118	0.0145	0.0136	0.0138	0.00148
Links/pg	8.33	8.33	9.11	9.11	9.10	8.80	0.43
Headings /pg	1.40	1.39	1.27	1.27	1.26	1.32	0.07
Largest page	426909	426909	426909	426909	426909	426909	0
Guiraud	70.4	70.6	71.1	71.3	71.4	71.0	0.4
Theor.vocab	2383	2388	2409	2421	2422	2405	18
Mean Dist.	4.5	4.4	4.5	4.6	4.6	4.5	0.08

Table 7: All key data from the five samples of the Sydsvenska Dagbladet web-site.

Since there are only five observations from the sites, we can only perform a tentative qualitative analysis of the data. Furthermore, some of the variables do not change much during the sample-period. Those variables are bytes, pages, types, links, bytes pr. page, types pr. page, links pr. page and theoretical vocabulary.

The sample clearly shows that something happened to the GP site between sample II and IV. Firstly, the number of link errors were 5, 4 and 3 in the previous samples, and suddenly increased to 85 and 91 in sample IV and V. What has happened, can be indicated by the fact that the mean distance also changed from 2.3 to 6.6. This indicated a major restructuring of the site transforming it from having a flat links-structure to a deeper one. The site had also obtained nine interactive forms from having no forms at in sample III. There has been a complete change in the sites outgoing links. The two most popular external links in sample IV were *www.realaudio.com* and *www.netscape.com* which occurred twenty times each. These links did not occur at all in any of the previous observations. In sample I-III the most popular links were www.westnet.com and www.sunet.se and they were referred about 25 times each. This indicates that there has been an overall change of the site's page and link-structure layout.

Although much weaker, SD also showed a change between sample III and IV, with increases in both the size of the site and in number of pages. Here the mean distance, however, remained virtually unchanged. In both GP and SD the tokens, types, Guiraud, and theoretical vocabulary showed that the types of texts did not change substantially. As an example, SD had an increase in tokens of around 9.5% over the period. This might not seem much, but the sample period was only 7 weeks, which roughly translates to 70% increase per year. It is not unrealistic to assume a steady growth, since the web site was started in August 1995.

Clustering Lexical Equality of All Newspaper Sites

At this level, all of the eight newspaper sites are analyzed in relation to each other by a cluster analysis on lexical equality of both types [see Figure 2], and tokens [see Figure 3]. The clustering algorithm visualizes the lexical equality percentages for each site in two-dimensions. There are no axes in the figure, only relation, e.g. the upper right corner is the least equal to the lower left corner. Each newspaper, except Svenska Dagbladet, is represented as five plots, each plot indicating a sample. Unfortunately, Svenska Dagbladet's sites blocked out our robot using the robots exclusion standard [Koster 1997] during the last two samples.

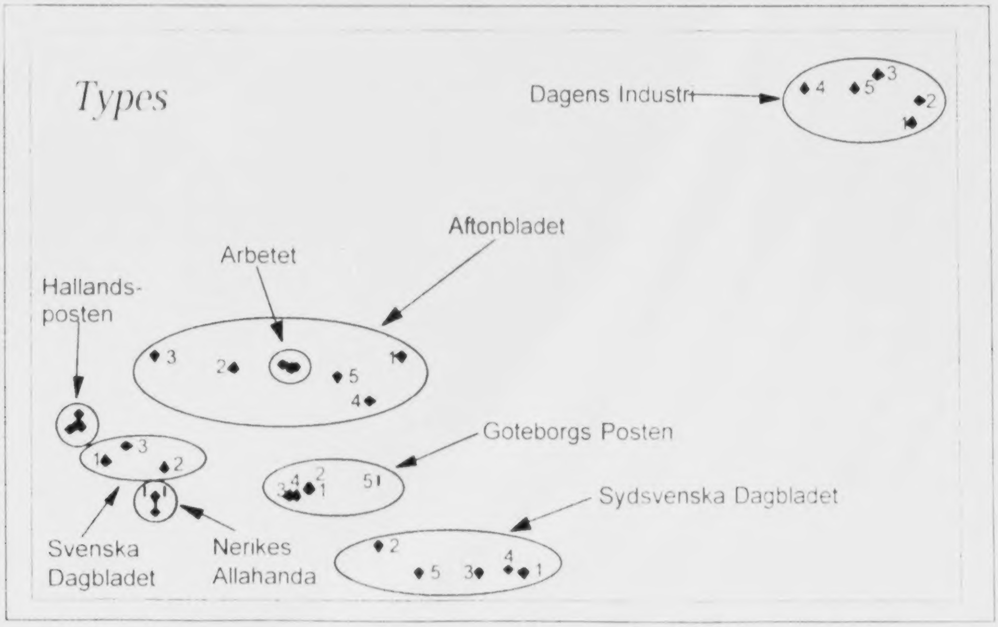


Figure 2: Clustering on lexical quality of types for the newspapers at five occasions.

The language on Arbetet, Hallandsposten, and Nerikes Allehanda changed very little over the period, which is not surprising because all of their other variables did not change very much either. All of the sites did, not surprisingly, change relatively little during the period, and therefore remain within a small region in the clustering. The variation in tokens and types within one particular site was less than the variation between sites. Dagens Industri is quite different from the others. The reason for this is probably that it is a financial newspaper that, due to the limited scope, uses a different language compared to the others. Arbetet and Aftonbladet are both associated with the Swedish Social Democratic Party. Although only the editorial in Aftonbladet has a distinct political flavor, it seems as though the language of the two newspapers is much the same. These two sites are quite equal in the token-based lexical quality test [Figure 3], but [Figure 2] shows an even greater similarity between the two when clustering types, i.e. comparing the two lists of distinct words. Göteborgs Posten and Svenska Dagbladet on the other hand are very different with the token-based method, but in [Figure 2] they seem to have many context carrying types that are the same.

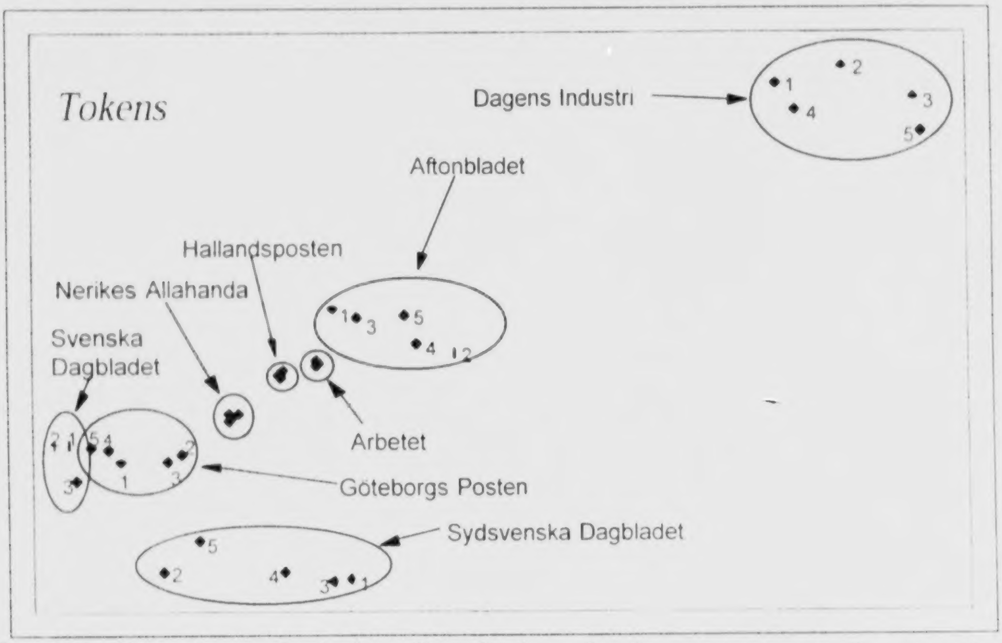


Figure 3: Clustering on lexical quality of tokens for the newspapers at five occasions.

6 Discussion

This paper has argued that we know very little about the fastest growing technology in the world. One of the ways we can learn more about the World Wide Web is to conduct quantitative surveys of the Web as a valuable supplement to research studying how people use it. This paper has investigated the question: *How can we apply semiautomatic quantitative measurement instruments to better understand changes to the contents of the Web?* The paper has investigated this question by presenting and discussing results from an experimental survey where a software robot collected and analyzed the contents of 82 Swedish Web-sites over a seven-week time-span from September 23 to November 4 1996. During the course of the experiment a couple of the sites banned robots from accessing data. Since we followed the ethical rules for robots, some of the time-series data were not complete but that is life, as we know it. In the near future we might see a replication of the old story of the nomads being blocked out by farmers building houses and drawing fences [Dahlbom and Janlert 1996].

We have shown two examples on how to obtain insight about changes to the contents of Web sites. Firstly, data from single sites has been analyzed. Here, we have conducted an analysis of the changes at the two Swedish newspapers Göteborgs Posten and Sydsvenska Dagbladet through 5 samples. Secondly, we compared the lexical equality of tokens and types in the 8 Swedish newspaper sites and showed the results graphically as plots.

The results showed the sampling instruments' ability to detect changes, but they also showed that a closer calibration of the instrument must be conducted during a longer time-span than seven weeks in order too increase the sensitivity to significant changes. As an example, the samples showed an increase in number of tokens of 9.5% over the seven weeks. If scaled to one year this amounts to around 70%. We know this is a large change, but compared to what? In general, a sampling period of only seven weeks most likely proved to be too short for obtaining results showing large variations.

The clustering graphs of lexical equality of tokens and types might not supply a strong statistical instrument for detecting significant changes but they proved to be good at graphically illustrating similarities and dynamics in Web sites.

In general, compiling frequency lists has provided us with much deeper material about the contents of the Web sites, compared to the analysis conducted by [Bray 1996]. On the other hand, this comes with the cost of a much smaller sample. Common readability formulas such as Coleman-Liau grade level and Bormuth grade level could be used to analyze the sampled texts. These indexes determine a readability grade level, based on characters per word and words per sentences and are therefore relatively easy to calculate. Word processors, such as, Microsoft Word uses these types of indexes in their grammar-checking facilities.

We attempted to overcome the problem of basing the sampling on sites by finding communities across several sites by looking at link structures. This turned out to be very complicated, and will be an item for further studies. Bray overcame the problem by only compiling frequency lists of external references for individual sites.

We have also tried to use search engines as slaves to detect how the sites are connected. Unfortunately, our population was too small to provide conclusive results. If a site between two sampling sessions split the site into two sites, then we will have to apply manual methods for detecting this. In general, the selection of target population it difficult to automate if the sample is to have a significant quality.

We did conduct initial analysis in order to detect statistical significant differences between the 82 Web sites grouped into the 6 sectors: companies registered on the Swedish stock-exchange; municipalities; newspapers; political parties and interest groups; government agencies; and TV- and radio stations. An analysis was also conducted of the same 82 sites grouped into the three groups: newspapers and others; media organizations and others; and private and public organizations. This analysis revealed a number of dependencies between, for example, the size of a site and the number of link errors. It also showed that the stock-exchange companies only have placed a brochure on the Web, whereas the media organizations have more dynamic sites. The analysis did, however, not show any further interesting results in terms of differences between sites from different sectors. This could mainly be caused by the very short sampling period, and by the absence of substantial contextual parameters. It might be so that in this stage of development of the Swedish Web sites, there is not much difference between different sectors' use of the Web as a publishing and advertising medium.

What are the possible implications of this research? Unlike many of the current research projects studying the Internet in general and the World Wide Web in particular, the effort reported here primarily has an analytical aim, i.e., to understand and describe the phenomenon in question, as opposed to a design-oriented perspective, for example, providing navigational support through visualizing and mapping information spaces. It might be argued that in computing or informatics research, an analytical perspective is always inferior to a constructive. Although this may be true, the lack of concepts and theories describing the nature of the World Wide Web can only emerge if we spend time studying and analyzing it. If we then ask ourselves what use we can make of a deeper knowledge of the Web, there is a large array of possibilities. In the following we will only provide pointers to some of the possibilities:

It could be interesting to analyze whether the language used on Web-based newspapers is similar or different to the language in the printed press. [Hagman and Ljungberg 1995] have conducted analyses of the contents of the printed press, thus making it feasible to make a comparison.

Increasingly companies are finding the Web an interesting place to market products. There are, however, only crude measures to guide a marketing strategy. The Web-sites might provide simple measures of the number of hits per day, but there is little help to find regarding where on the Web a company should target its advertising efforts. If, for example, Volvo would like to know where on the Web to market their newest car model, they could be interested in knowing where cars are intensively discussed and where they are not.

The relationships between the contents of the Web and the perception of it, is also an interesting avenue for further research. For example, companies must choose an appropriate level of interactivity when presenting themselves on the Web. It is a choice between providing a Web-based brochure or building a heavily interactive system defining new and interesting ways of bringing the customer closer to the company. It is, therefore, interesting to investigate whether interactivity in companies' Web pages can be viewed as a measure of customer orientation. Answering questions like this demands a combined approach where the contents of selected Web-sites must be analyzed using qualitative methods.

Statistical surveys of population groups operate on a small sample of the entire population. Random sampling is, therefore, an important issue. One way of sampling randomly is to use the telephone directory, since most households have a telephone. Assuming that we are just beginning to move to the net, with the Web user population moving towards an average sample of the population of Western industrialized countries, how can we then conduct random statistical sampling of the population? In order to do so, we must have concepts and theories describing the Web from both a qualitative and quantitative perspective.

The Dow Jones Industrial Average (www.dowjones.com) of 12 leading North American companies appeared first in the precursor of The Wall Street Journal, May 26, 1896. The index provides a means of gauging the stock market. Would it not be a good idea to compile a similar index for the Web? That would allow us to register when: the use of Sunsoft Java goes up and Microsoft Active X goes down; the use of frames is as popular a navigational aid as tables; or client-side clickable maps have entirely replaced CGI-scripted maps. Existing search engines could be applied to accomplish this. Using others' search engines as slaves has been done by MetaCrawler and Ahoy with very impressive results.

We hope that the examples above have provided a perspective to the discussion of why it is crucial to conduct a quantity survey of the Web. Within the Internet Project we will continue to develop interesting measures, test these in data-collection experiments and translate the results into concepts and theories describing the Web. Because the Web is a highly dynamic and interactive information space, we must apply state-of-the-art computational power to study its nature and progress.

Acknowledgements

This research is partly funded by the Swedish Transport and Communications Research Board (Kommunikationsforskningsberedningen) grants for the Internet Project. A big thanks to Leif Grönquist for assistance on computational linguistics and to Michael Mandahl for coming up with some bright ideas. Also a huge thanks for all the constructive comments to all our other colleagues in the Internet Project.

References

[Berners-Lee et al. 1994] Berners-Lee, T., Cailiau, R., Luotonen, A., Nielsen, H. F., & Secret, A. (1994). The World-Wide Web. *Communications of the ACM* 37 (8), 76-82.

[Barry and Batty 1994] Barry, B. & Batty, M. (1994). The Electronic Frontier - Exploring and mapping cyberspace. *Futures* 26 (24), 699-712.

[Blundon 1996] Blundon, W. (1996). Off The Charts : The Internet 1996. *Internet World (12)*.

(Also at the URL: http://www.iw.com/1996/12/charts.html).

[Bray 1996] Bray, T. (1996). Measuring the Web. *In proceedings of the 5th International World Wide Web Conference*. Paris, France.

[COMNET 1996]. CommerceNet / Nielsen Internet Demographic (1996). (At the URL: http://www.commerce.net/nielsen/).

[Dahlbom and Janlert 1996] Dahlbom, B. & Janlert, L.-E. (1996). Nomaden im Cyberspace. *Bertelsmann Briefe*, Heft 135, Frühling/Sommer.

[Dodge 1997] Dodge, M. (1997). The Geography of Cyberspace.

(At the URL: http://www.geog.ucl.ac.uk/casa/martin/geography_of_cyberspace.html).

[Dodge 1996] Dodge, M. (1996). Mapping the World Wide Web. *GIS Europe* 5 (9), 22-24.

[Drew and Hendley 1995] Drew, N. S. & Hendley, R. J. (1995). Visualising Complex Interacting Systems. *In ACM SIGCHI '95 Proceedings of the Conference on Human Factors in Computing Systems*. Denver, Colorado, USA.

[Eichmann 1994] Eichmann, D. (1994). Ethical Web Agents. *In the proceedings of the 2nd International World Wide Web Conference*. Chicago, USA.

[Fielding 1994] Fielding, R. T. (1994). Maintaining Distributed Hypertext Infostructures: Welcome to MOMspider's Web. *In the proceedings of the 1st International World-Wide Web Conference*. Geneva, Switzerland.

[Girardin 1996] Girardin, L. (1996). Mapping the virtual geography of the World-Wide Web. *In proceedings of the 5th International World Wide Web Conference*. Paris, France.

[Hagman and Ljungberg 1995] Hagman, J. & Ljungberg, J. (1995). Brute Facts versus Institutional Facts of Language as a Foundation for IR. *In the proceedings of the 5th European-Japanese Seminar on Knowledge Bases and Information Modeling*. Sapporo, Japan.

[Hoffman et al. 1996] Hoffman, D. L. , Kalsbeek, W. D. & Novak, T. P. (1996). Internet and Web use in the U.S. *Communications of the ACM* 39 (12), 106-108.

[Jain and Dubes 1988] Jain, A. K. & Dubes R. C. (1988). *Algorithms for Clustering Data*. Prentice-Hall. NJ: Englewood Cliffs.

[Kelly and Wolf 1997] Kelly, K. & Wolf G. (1997). Push! Kiss Your Browser Goodbye: The Radical Future of Media beyond the Web. *Wired UK 3* 69–81.

[Koster 1995] Koster, M. (1995). Robots in the Web: threat or treat?. Technical report. ConneXions 9 (4)

[Koster 1997] Koster, M. (1997). The Robots Exclusion Protocol.

(At the URL: http://info.webcrawler.com/mak/projects/robots/exclusion.html).

[Mitchell 1995] Mitchell. W.J. (1995). *City of Bits: Space, Place and the Infobahn. The MIT Press*, USA.

[Monty Python 1980] Monty Python. (1980). The Bookstore. *On Monty Python's Contractual Obligation Album*. Charisma 1152 (U. K.)/Arista AL 9536 (U. S.). (Also at the URL: http://www.nettap.com/~drc/python/BOOKSHOP.html).

[Mukherjea et al. 1995] Mukherjea, J. & Foley, J. D. & Hudson, S. (1995). Visualizing Complex hypermedia networks through multiple hierarchical views. *In ACM SIGCHI '95 Proceedings of the Conference on Human Factors in Computing Systems*. Denver, Colorado, USA.

[Mukherjea and Foley 1995] Mukherjea, S. & Foley, J. D. (1995). Visualizing the World-Wide Web with the Navigational View Builder". *Computer Networks and ISDN System, Special Issue on the 3rd International Conference on the World-Wide Web '95, April 1995, Darmstadt, Germany*.

[Negroponte 1995] Negroponte, N. (1995). *Being Digital*. London: Coronet Books. Hodder and Stoughton.

[Olsen et al. 1993] Olsen, K. A. & Korfhage, R. R. & Sochats, K. M. & Spring, M. B. & Williams, J. G. (1993). Visualization of a Document Collection with Implicit and Explicit Links - The Vibe System. *Scandinavian Journal of Information Systems* 5, 79-95.

[Pitkow and Kehoe 1996] Pitkow, J. E. & Kehoe, C. M. (1996). Emerging trends in the WWW User Population. *Communications of the ACM* 39 (6), 106-108.

[Rickard 1995] Rickard, J. (1995). The Internet by the Numbers - 9.1 Million Users Can't Be Wrong. *Boardwatch Magazine* 9 (12).

[SCB 1996] Statistisk Årsbok 96 (1996). Statiska Centralbyrån. Sweden.

[Tesitelová 1992] Tesitelová, T. (1992). *Quantitative linguistics*. Amsterdam: John Benjamins Publishing Company.

[Voionmaa 1993] Voionmaa, K. (1993). *On the Semantics of Adult Verb Acquisition*. Ph. D. Theses. Department of Linguistics. Göteborg University.

'Agentizing' the Internet – the Practical Design of a Networked Mobile Agent System

Benjamin Falchuk, Ahmed Karmouch
Department of Electrical & Computer Engineering, University of Ottawa
161 Louis Pasteur, Ottawa, ON, CANADA, K1N 6N5, Tel: (613) 562-5800x6203 Fax: (613) 562-5175
E-mail: ben@sol.genie.uottawa.ca, karmouch@elg.uottawa.ca Web: http://deneb.genie.uottawa.ca

Abstract: Currently the World Wide Web dominates the Internet scene. It is, however, characterized by its user-driven mode in the sense that nothing happens upon the web browser's screen until the user makes a mouse selection to download a new page, start an applet, etc. The mobile agent-model offers a considerably more flexible approach to information discovery and retrieval. When viewed as an assistant to a human user it is clear that an agent's asynchronous, collaborative, and mobile nature provides savings on both the time and bandwidth dimensions. This paper introduces mobile agents, describes the main issues involved in creating an agent-based system on the Internet, and then describes *AgentSys*, a prototype mobile agent system. *AgentSys* allows mobile agents to roam from execution environment to execution environment, collect and negotiate for information – including "talking" to HTML documents – and return to the user with results. The protocols used in *AgentSys* are novel extensions to existent communications protocols and drafts. This work suggests that mobile agent environments will soon become the second most important Internet servers next to, and in conjunction with, those supporting the HTTP protocol.

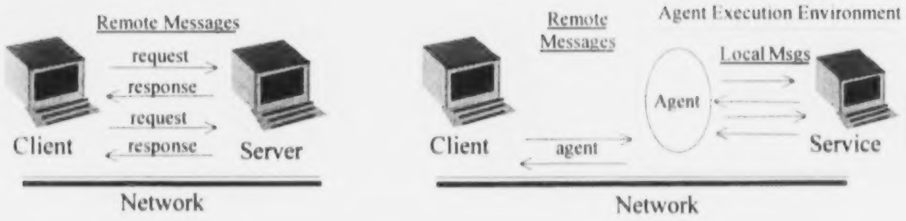
Introduction

The global Internet and the WWW, HTTP, and HTML [Berners-Lee and Connolly 95] standards have created an information revolution of sorts. It is now possible to find a huge amount of information on-line, ranging from personal home pages to medical databases. *Netscape Navigator* and *Internet Explorer* have become the two de facto modes of exploring and searching the web. As client-server (C-S) tools however, these programs put the onus on the user to continually follow hyperlinks to find desired information. Although search tools such as *Lycos* [Lycos 97] and *WebCrawler* [Webcrawler 97] exist, it is still time-consuming, bandwidth wasting, and tedious to find information. The mobile agent paradigm can help to alleviate these problems.

Mobile Internet Agents

The mobile agent-model extends the basic concepts of Postscript and Remote Evaluation (REV). Four main concepts are integral: (i) agents - programs or scripts, compiled or not, that represent the user, (ii) agent execution environments - virtual machines in which agents can run, (iii) resources - CPU cycles, memory, disk space, services (including databases, WWW data, etc.), and (iv) protocols - timing, syntax, and semantics that allow agents to intercommunicate. Functionally, the processing that would have occurred between the user and the client-GUI that represents the remote service is encapsulated in the agent and physically co-located with the service.

A general and informal definition of a mobile agent is as follows: *a program that is able to migrate from node to node on a network under its own control for the purpose of completing a task specified by a user. The agent chooses when and to where it will migrate and may interrupt its own execution and continue elsewhere on the network.* Note that Web Spiders, Robots, and *Lycos* are *not* mobile agents by this definition (see [Cheong 96]). [Fig. 1] illustrates a comparison to the client-server model. Mobile agents do not require network connectivity with remote services in order to interact with them and a network connections are used for one-shot transmissions of data (the agent and possibly its *state* and *cargo* [Ford and Karmouch 97]) and then closed. Results in the form of data do not necessarily return to the user using the same communications trajectory, if indeed results are expected at the node. Alternatively, the agent may send itself to another intermediate node and take its partial results with it. Results are delivered back to the user whose address the agent knows. In general we can say that the mobile agent-model offers the following advantages over C-S: (i) uses less bandwidth by filtering out irrelevant data (based on user profiles and preferences) at the *remote* site *before* the data is sent back, (ii) ongoing processing does not require ongoing



(a) the client-server model

(b) the agent-model

Figure 1: The client-server model (e.g. browsing the WWW) versus the mobile agent-model

connectivity, (iii) saves computing cycles at the user's computer, (iv) it is more efficient as the processing moves closer to the data, and (v) frees the user to log out or migrate since the agent's life is independent of the user's session. The remainder of this paper is structured as follows: we first discuss our mobile agent platform architecture, we then describe inter-agent communications, and finally conclusions are drawn.

A Prototype Agent Architecture – the New Internet Servers

In a mobile agent system there is a network (Internet or LAN), human users with the freedom to program and insert agents onto the network, and service providers offering various services to the human and mobile agent community. Examples of services are WWW data stores, video or music-on-demand, and agent meeting places [White 94]. Mobile agents roam between agent execution environments (AEE's) on the network in search of media that satisfy a user request. The complexity of mobile agent operation requires elaborate "servers" and while HTTP servers are relatively simple, agent "servers" must support all aspects of mobile code operation. Agents are more than just messages – they are *programs* that require memory and disk-space, engage in conversations, carry user-information and cargo, have goals, and have some degree of autonomy.

In our mobile agent system, the user specifies his high-level request with the *AgenTask* protocol that allows under-specification. Agents are transferred by local transfer entities using the *AgenTransfer* protocol that runs over TCP/IP. Arriving agents are accepted from the socket by transfer entities, unpacked and passed on to a facilitator entity that creates a "virtual machine" environment for the incoming agent. Facilitators manage local resources and serve as the only gateways through which incoming agents may access these resources. Using *AgenTalk* messages, agents are free to communicate with other mobile agents (local or remote), request services from facilitators, or migrate. *AgenTalk* is discussed later, however *AgenTask* and *AgenTransfer* are not discussed further in this paper.

The Agent Execution Environment (AEE)

The AEE [Fig. 2] is divided into several functionally distinct modules that inter-communicate and share responsibility. Almost all modules use local memory or disk space to create caches, tables, or databases for the purpose of storing persistent data relating to their functional specification. Physically these data-stores are not considered as parts of the AEE. The facilitator is the coordinator of the AEE in the sense that (i) it must occasionally ensure that all other modules are operational, and (ii) it serves as the gateway to all local resources and delivers messages to mobile agents that it hosts.

Facilitator Module – Mobile agents request services and communicate with other agents through this *gateway*. The facilitator has a well-known port number through which it can be reached. It is the role of the facilitator to occasionally ensure that the other modules are operational. The *AgenTalk* module implements the protocols necessary for inter-agent messaging and conversations and is based loosely on KQML [Finin et al. 94].

Agent Data Manager – Incoming mobile agents may arrive with, or subsequently acquire, media. When an agent arrives and is followed by media it is given a *handler* to the media which are physically placed on tertiary storage. The actual location and schema of the stored media is transparent to the mobile agent. The salient point is that when the agent subsequently migrates to another node, the data manager maps the agent identifier to its media and then gives them to the transfer module. If the storage area that the data manager handles becomes too full, a message is sent to the transfer module and incoming agents with cargo exceeding a fixed number of bytes may be rejected since the persistence of their cargo cannot be guaranteed. It is also the case that the acquisition of a media may be rejected if that media will overflow the agent's individual storage space restrictions.

Local Resource Monitor – Regardless whether the node hosting the AEE is a stand-alone server or a workstation, other processes may be active, including user-sessions, print jobs, or other mobile agents. The local resource manager's role is to collect and monitor information on CPU load, disk-space, and other critical aspects. This module sends interrupts to the facilitator module when, for instance, high CPU usage by other more critical processes must pre-empt the interpretation of a mobile agent.

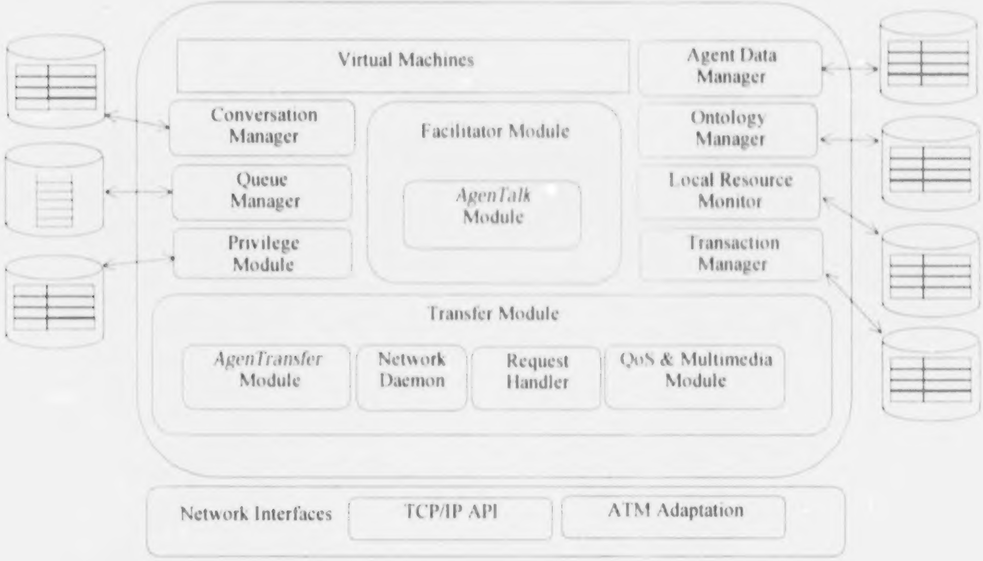


Figure 2: The architecture of an agent execution environment – the next Internet servers

Transaction Manager – Mobile agents use *AgenTalk* to request media - these requests are mapped to local storage based on both the primitive and the ontology. To allow concurrent interleaved access to local data-stores, the transaction manager maintains the integrity of requests and data, forms sub-transactions if necessary, and writes checkpoints and recovery information for agent transactions.

Ontology Manager – Individual ontology specifications and their versions are stored on disk and must be managed by the ontology manager. A remotely operating agent may requirean ontology (e.g. by *name* and *version*) be sent to it – in this case the request is received by the transfer module which asks the ontology manager for the data. The data is sent to the transfer module and then over the network to the remote transfer module. Ontologies may be stored using any representational syntax that is interpretable by the facilitator (e.g. *Ontolingua* [Gruber 93])

Conversation Manager – While in conversation, agents send messages to each other through the facilitator module. These messages are passed to the conversation manager whose role it is to validate the messages in the context of the conversation, and then log them. The conversation manager has a knowledge base that describes the conversation protocols which it uses to check messages for integrity and to make logs in the form of relations (including timestamps). When the agent's user requests a detailed record of agent activity the facilitator requests the appropriate logs. This module also stores extensions to the *AgenTalk* protocol when necessary.

Queue Manager – When incoming mobile agents cannot be executed immediately they are placed on a waiting list. This occurs when the Local Resource Manager deems that the CPU is too heavily loaded (if there is insufficient storage for the agent's cargo then the agent is rejected regardless of CPU load). The queue manager reports the current queue contents to the facilitator. This data includes the agent identifier (for local purposes), the agent's task type, human-user identification, and a handler into the Data Manager's resources for that agent. This allows the facilitator to answer queries regarding the agents in the queue as well as active ones (e.g. a mobile agent can ask the facilitator, "Are there any other agents, active or queued, with a similar task-type as mine?")

Privilege Module - Incoming mobile agents must be authenticated. The local AEE maintains a list of human users and their authentication. Furthermore, agents acquire privileges to use certain resources. The privilege module stores records that indicate *what* the agent may access and stores securely the associated passwords.

Virtual Machines - Regardless how agents are implemented (e.g. declaratively or procedurally), they require a virtual machine in which to run. This is typically an interpreter (e.g. LISP, Java etc.), forked by the facilitator to processes the agent's executable code line by line. Depending on the language, agents may be able to create local variables, open files for reading, etc. *AgentSys* is Java-based.

Transfer Module - The transfer module is a set of sub-modules that handle the transfer and reception of mobile agents between nodes. Each node has a Network Daemon that listens for, reads, and sends messages to the *AgenTransfer* module that implements the transfer protocol and its syntax. The Request Handler examines the primitive (e.g DISPATCH [IBM 97]) and the parameters and then does one of many things. If the incoming agent has cargo to follow, a local identification is made for that agent and given to the Agent Data Manager. When each piece of cargo subsequently arrives, it is given to the Data Manager who stores is appropriately. The Request Handler thus implements a type of "session" in which cargo arrives piece by piece. The Request Handler passes validated incoming agents to the Facilitator for interpretation. The Quality of Service (QoS) and Multimedia Module allows for negotiation of bit rate and priority. Agents marked as high priority are less likely to be placed in the Queue (at some financial cost to the agent). Agents that negotiate for higher bit rates or lower latency are sent through an ATM adaptation layer (AAL) for segmentation into cells, and then onto the fiber-optic network.

AgenTalk - An Inter-Agent Communication Language

A mobile agent is like a businessperson going from country to country. Not only does it need the correct passports and authorization to enter each country, it also needs the intelligence to speak in the *native* language or in some "universal" language. In this section we introduce the key elements required to allow a community of possibly heterogeneous agents to talk about and exchange data and messages, help each other, and migrate from node to node. Existing drafts and standards for these purposes include KQML [Finin et al. 94] and ATP/0.1 [IBM 97].

Data, agents, networks and the services that they provide are complex and multi-faceted. It is clear that if the goal is to have an agent system in which mobile agents roam a network where services are offered, transactions are made, and results gained, then it is crucial to have what is referred to as *an ontology*. As defined in [Gruber 93], an ontology is, "...a common vocabulary in which shared knowledge is represented [that]...associates names of entities...with human-readable text describing what the names are meant to denote, and the formal axioms that constrain the interpretation and well-formed use of these terms." An ontology removes ambiguities that may occur between communicating parties, including human-agent and agent-agent modes. For example, by agreeing that the term `agent-content.agent-task-type` refers to an agent's task description and by limiting it to a 64-byte *string*, this data can be exchanged and understood without confusion. *AgentSys* uses a set of application ontologies that describe data at four critical levels: *agents, services, documents,* and *networks.* The complete specification of the hierarchy is beyond the scope of this paper.

Conversational Modes - "Talking" to Web Documents

AgentSys uses a small set of speech-acts that are the building blocks of inter-agent conversations. A speech-act consists of a performative and content. For example, *tell (agent1, k)* is a speech-act, *tell* is the performative and *agent1, k* is the content. Speech-acts may have different meanings in different contexts. For instance, an *ACK* message in response to a counter proposal may indicate acceptance, whereas an *ACK* in response to a document component may imply "send more information". The main goals of the conversation policies established here are to: (i) allow agents to authenticate one another, (ii) allow for the inter-agent exchange of: protocols, documents, user-information, and other queries, (iii) enable negotiations to occur between agents and facilitators as well as between two user-agents, and (iv) allow agents to help each other solve tasks in a number of ways. [Tab. 1] illustrates the primitive verbs and the conversation policies.

Primitive verbs (see [Finin et al. 94]) in the *AgentSys* system appear within conversation policies, and as shown in [Fig. 3], state diagrams can represent those policies. In these diagrams, the dashed circle is the start state and the dark circles are end states. The other circles are intermediary states with the receiving entity shown in numeric form. The arrows are transitions and the text is the verb in the conversation that invokes that transition. Since many conversations require the transfer of possibly sensitive data a *tell_auth* or an *ask_for_auth* conversation precedes almost all other interactions.

Primitive verbs	*Conversation policies*	*Functionality of policy*
ACK	Exchange_docs	Each agent provides a document to the other
Decline	Help_with_doc	One agent suggests media that "fits" into another's document
Reject	Talk_to_webdoc	An agent asks for representative data from a WWW document
Offer	Talk_to_synchdoc	An agent asks for representative data from a synchronized document
Suggest	Query_user_info	Ask for information about another agent's human user (e.g. E-mail)
Command	Ask_for_protocol	Exchange a protocol representation (in state, next-states format)
Inform	Provide_agent	Ask for an agent that might be able to help with the current task
Accept	Negotiate_for	Engage in an iterative, interactive negotiation for data
Propose	Ask_for_auth	Two agents authenticate each other
Ctr_propose	Tell_auth	One agent demands authentication from another
	Help_coop	An iterative collaboration between two agents
	Help_one	One agent helps another
	Suggest_server	A remote facilitator that might be helpful is suggested
	Provide_connection	A connection is provided to a remote agent server
	Acquire	Associate the named media with the agent – e.g. a "purchase"
	Receive_msg	Receive the next message queued for the agent

Table 1: Verbs and conversations specified in *AgenTalk*

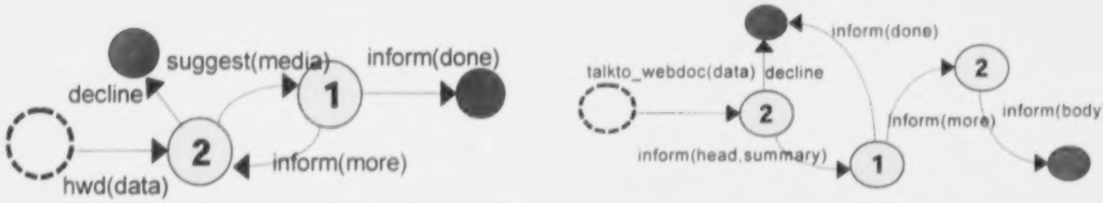


Figure 3: Conversations between agents are modeled as finite state diagrams. At each stage a message is sent and the response comes from a finite set of messages. (hwd = help_with_document)

The *talkto_webdoc* conversation policy allows a mobile agent to ask for the critical portions of an HTML document. Once a handler to the document name has been acquired, the mobile agent issues the *talkto_webdoc* message, naming the document and other parameters, and then waits. The facilitator agent accepts the message, retrieves the header information and passes it back to the mobile agent. The mobile agent receives the response, does some arbitrary processing, and then asks for the body information using an *inform* message. The body data is then returned to the agent by the facilitator. If after this sequence the mobile agent wishes to add this document to its cargo, an *acquire* sequence is started.

Conclusions

Among the stumbling blocks with Internet agent-systems is that they require a large-scale acceptance and adoption of the protocols. Our experiments with *AgentSys* have re-emphasized the main issues with 'agentizing' the Internet:

- The WWW and the HTTP protocols are not suitable to support full-blown mobile agent operation.
- A new set of standard protocols for mobile agents must be developed and adopted by the Internet community. Proprietary agents and standards have emerged from Telescript [White 94] and IBM [IBM 97], as have agent standards groups, including the Agent Society [Agent 97] and FIPA [FIPA 97].
- Security is a stumbling block for commercial adoption of the agent-model. Most users will not host mobile agents on their system until they are sure they can defend themselves from malicious and mischievous agents.
- All access to local resources (such as web data and databases) by mobile agents should be through a facilitator agent so that incoming mobile agents do not have *direct* access to resources. WWW resources remain crucial.

The *AgentSys* mobile agent system is a prototype system built upon a networked Pentium 199MHz Win95/NT platform and some test sites on the Internet. Our agents are transferred using an extended IBM ATP/0.1 protocol [IBM97] over Ethernet or ATM (Madge^{TM} Collage 120) networks. Media resources on our test-bed include ObjectStore^{TM} object-oriented databases, and WWW pages. The system has been developed using both Java [SUN 94] and TCL [Ousterhout 94]. [Fig. 4] illustrates some of the GUI's that allow behavior specification.

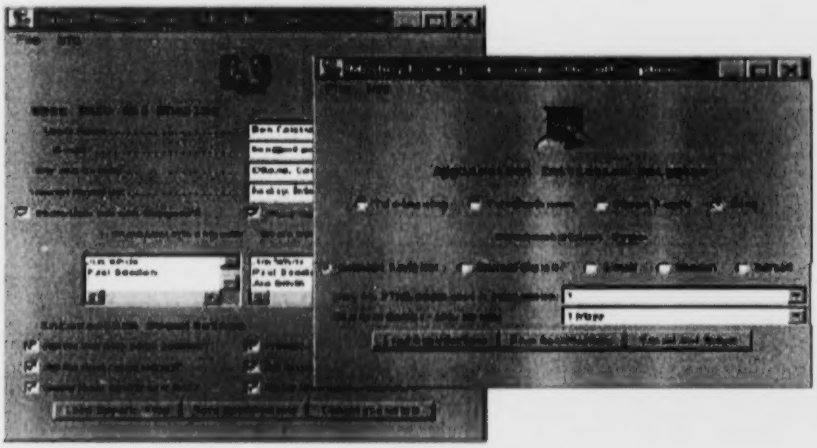


Figure 4: Specifying aspects of mobile agent behavior using Java-based interfaces.

Despite the stumbling blocks that hinder the widespread adoption of mobile agent systems, we feel that our experiments have proven that the agent model is practical, saves time, and reduces bandwidth usage on the Internet. By moving code to data mobile agents can become effective assistants to humans for tasks that can be automated or are tedious or repetitive.

References

[Agent 97] The Agent Society, http://www.agent.org/

[Berners-Lee and Connolly 95] Berners-Lee T., Connolly D. (1995). Hypertext Markup Language – 2.0, IETF HTML Working Group, http://www.cs.tu-berlin.de/~jutta/ht/draft-ieft-html-spec-01.html

[Cheong 96] Cheong F.C., *Internet Agents: Spiders, Wanderers, Brokers, and Bots.* (1996). New Riders Publishing, Indianapolis

[Finin et al. 94] Finin T., McKay D., McEntire R. (1994). KQML as an Agent Communication Language, *Proc. Third Int. Conf. On Information and Knowledge Management*, ACM Press

[FIPA 97] Federation for Intelligent Physical Agents, *http://www.cselt.stet.it/ufv/leonardo/fipa.htm*

[Ford and Karmouch 97] Ford B., Karmouch A. (1997). An Architectural Model for Mobile Agent-based Multimedia Applications, *Proc. of Canadian Conference on Broadband Research '97*, Ottawa

[Gruber 93] Gruber T. (1993). A Translation Approach to Portable Ontology Specifications, Knowledge Systems Lab, Stanford University, Computer Science Department, Technical Report KSL92-71

[IBM 97] Aglets Workbench, Agent Transfer Protocol ATP/0.1, http://www.trl.ibm.co.jp/aglets

[Lycos 97] Lycos, http://www.lycos.com

[Ousterhout 94] Ousterhout J.K., ed. (1994). *Tcl and the Tk Toolkit*, Addison-Wesley Publishing Company, New York

[SUN 94] Sun Microsystems (1994). The Java Language: A White Paper, Sun Microsystems White Paper

[Webcrawler 97] Webcrawler, http://www.webcrawler.com

[White 94] White J.E. (1994). Telescript Technology: The Foundation for the Electronic Marketplace, General Magic White Paper

Acknowledgments

This work has been supported partly by the Natural Science and Engineering Research Council (NSERC) and the Telecommunications Research Institute of Ontario (TRIO).

BEST COPY AVAILABLE

Publishing Troubleshooting Expertise on the World-Wide Web: A Case Study

Robert Farrell
Web Applications Group, Bellcore, USA
E-mail: farrell@bellcore.com

Karen Kukich
Educational Testing Service, USA

Frank T. Jones III
Service Management Platform Group, Bellcore, USA

Abstract: The expertise needed for troubleshooting large commercial software systems is often distributed throughout an organization. The World-Wide Web provides an ideal medium to collect and redistribute this expertise to users on diverse computer platforms. This paper offers a case study of our efforts at gathering, structuring, and distributing the information needed to troubleshoot environment-related problems with a large telecommunications software system. Our troubleshooting expert system provides Web access to a knowledge base of over 450 cases. Individual cases can be accessed using full-text searching, browsing, or guided problem-solving.

Background

Bellcore provides software and consulting services to the global telecommunications market, including many of the operation support systems that run the networks to support local phone service. Our company offers a new scaleable suite of operations support systems that use a client-server architecture, open interfaces, and shared corporate data to support video and telephony services. These products support order entry and tracking, service activation, network inventory, work force management, and other critical functions for telephone and cable companies.

The various clients and servers that are part of this new telecommunications operations product suite communicate using the Open Software Foundation's Distributed Computing Environment and share a common graphical user interface. In this distributed environment, when software components fail it is often a difficult troubleshooting task to identify the root cause of a problem and a develop a repeatable resolution procedure. Troubleshooting must often be performed by highly-skilled software developers, a costly proposition. Furthermore, these developers require extensive training and their expertise is lost if they move to another project or leave the company.

Data from customer service departments suggest that most of the problems reported by customers do not originate from software application bugs, but are the result of customer-specific environment problems or middleware. These problems may arise during installation, configuration, or ongoing management of the application. We quickly determined that automated troubleshooting [Hamscher et al. 1992] of these problems based on a complete model of the software was too difficult. However, if we could capture the expertise of developers, testers, and field engineers at diagnosing these problems and distribute that knowledge to all of our customers, we could empower our customers' system administrators to solve many of their own problems and reduce calls to our help desk.

A Manual Approach to Troubleshooting

Early in 1996, the organization responsible for developing our new product suite decided to proactively capture and document the most likely environment-sensitive problems. They formed a troubleshooting task force charged with compiling a list of potential points of failure for each product. Each point of failure was represented as a case, including

the symptoms of that failure as observed by an end user or system administrator,
the diagnostic procedures to follow to confirm the problem, and
the restoration procedures to follow to resolve the problem.

All cases were gathered into a troubleshooting manual. The development organization also produced a second document, an error code manual. This manual lists error codes for each product along with their associated textual messages and resolution procedures. Together these two documents form a central repository of troubleshooting knowledge. However, the documentation has the following disadvantages:

Updates - The troubleshooting information could only be updated on the publishing cycle of the documentation, roughly once every four months. Problems not anticipated by developers would be rediscovered at each customer installation.

Consistency - Developers were not consistent in how they coded information for documentation. For example, diagnostic information applicable across products was often repeated in slightly different language.

Information access and usage - The paper manuals provided little help in finding the right failure point, given a particular symptom. A "roadmap" was developed, but that provided only high-level guidance. Users had to learn troubleshooting on their own and leaf through the paper manual to locate relevant cases.

An On-line Troubleshooting Expert

In mid-1996, we started to investigate how to encode troubleshooting information in a more consistent form, accessible to everyone who develops, tests, and deploys the product line. We wanted a framework that would support entering symptom and cause information, but would not require coding of complex rules. We needed a delivery method that would help users navigate swiftly through the knowledge base (KB) to locate the root cause of a problem and discover its resolution. This required a paradigm that would support searching over the entire KB combined with diagnostic question-answering to help narrow the focus to a probable root cause for a problem.

Motivations for Web Delivery

We anticipated that the number and diversity of users wanting access to the KB would grow as we moved from deployment within Bellcore to our customer sites. Access through a Web browser would allows us to quickly deploy to multiple locations, permit users to access the KB from diverse workstation platforms, maintain some consistency in appearance between the on-line presentation and the manual documentation, and support a simple and familiar user interface.

System Features

Late in 1996, we started creating the current system called TroubleShooting Expert (TS/E). TS/E includes the following troubleshooting aids:

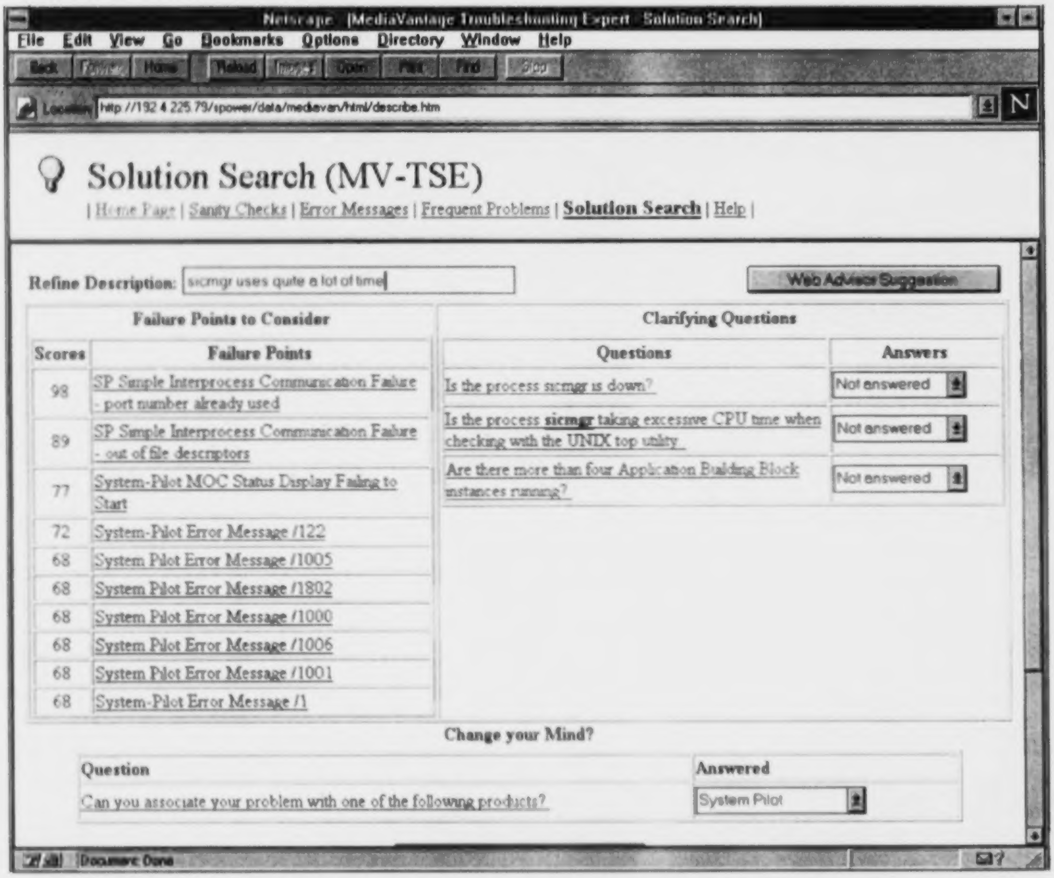
Sanity Checks - Diagnostic tests to verify the integrity of the system. Failing a test leads to a listing of one or more probably causes.

Error Message Lookup - A list of error codes and messages. Selecting a code leads directly to one or more resolution procedures.

Frequent Problems - A list of problems that arise frequently. Selecting a problem leads to a listing of one or more probable causes.

Solution Search - A natural language search restricted by product category. Submitting a query results in a rank-ordered list of probable causes.

Figure 1 shows a screen shot from TS/E during a solution search. The user entered the query: "sicmgr uses quite a lot of time". TS/E gives a ranked list of cases (failure points) in the left column together with a list of diagnostic questions in the right column. Answering any of the questions will re-rank and eliminate some cases, thus narrowing the problem focus. Selecting a case leads to its problem resolution page.


Netscape - [MediaVantage Troubleshooting Expert - Solution Search]
File
Edit
View
Go
Bookmarks
Options
Directory
Window
Help
Back
Home
Reload
Find
Stop
Location:
N
| Home Page | Sanity Checks | Error Messages | Frequent Problems | Solution Search | Help |
Web Advisor Suggestion
Document: Done

http://192.4.225.79/spower/data/medievan/html/describe.htm

Solution Search (MV-TSE)

Refine Description: sicmgr uses quite a lot of time

Failure Points to Consider

Scores	Failure Points
98	SP Simple Interprocess Communication Failure - port number already used
89	SP Simple Interprocess Communication Failure - out of file descriptors
77	System-Pilot MOC Status Display Failing to Start
72	System-Pilot Error Message /122
68	System Pilot Error Message /1005
68	System Pilot Error Message /1802
68	System Pilot Error Message /1000
68	System Pilot Error Message /1006
68	System Pilot Error Message /1001
68	System-Pilot Error Message /1

Clarifying Questions

Questions	Answers
Is the process sicmgr is down?	Not answered
Is the process **sicmgr** taking excessive CPU time when checking with the UNIX top utility	Not answered
Are there more than four Application Building Block instances running?	Not answered

Change your Mind?

Question	Answered
Can you associate your problem with one of the following products?	System Pilot

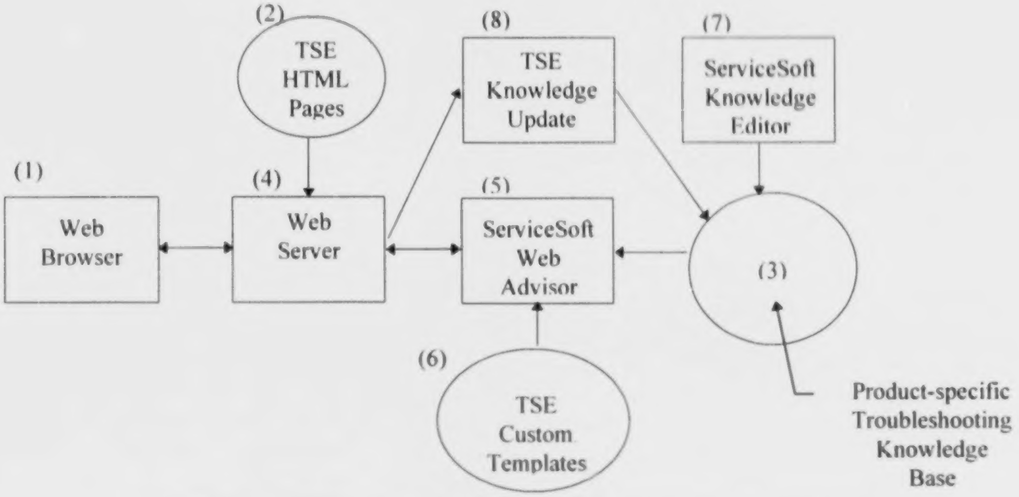
Figure 1: TS/E lists probable Failure Points during a Solution Search

System Architecture

TS/E uses an inference engine from ServiceSoft Corporation to rank-order cases based on the combination of word matching and answers to diagnostic questions. ServiceSoft provides

- an easy-to-use knowledge editing environment (Knowledge Editor),
- a diagnostic engine that combines natural language search with case-based reasoning (Web Advisor™),
- an object-oriented KB schema that roughly matches our point-of-failure cases, and
- a customizable Web interface based on templates.

Figure 2 depicts our system architecture based on ServiceSoft's design. Web Advisor actually consists of a CGI script and an intermediate data server [Varela et al. 1995] that keeps the connection to the KB open for improved efficiency. Templates are coded in a proprietary HTML extension that provides some flexibility in ordering and presenting the information retrieved from the knowledge base.



Users access TSE through a standard Web browser (1), starting with TSE-specific HTML pages (2). The knowledge base (3) is accessed and formatted into dynamic Web pages for the Web Server (4) by ServiceSoft's Web Advisor (5) according to the TSE's Custom Templates (6). Cases can be entered using ServiceSoft's Knowledge Editor (7) or can be entered in a special form that is processed by TSE's Knowledge Update facility (8).

Figure 2: TS/E Architecture

Populating the Troubleshooting Knowledge Base

A complete, up-to-date, and accurate KB is critical to the success of any knowledge-based system. While ServiceSoft's Knowledge Editor provided a convenient environment to edit an existing KB, we needed a method for converting the existing documentation into an initial KB. We also needed a way to update the KB as subject matter experts and users discover unanticipated problems. This section describes how we converted, imported, and restructured the documentation to create point-of-failure cases, and how we enhanced TS/E to capture and format new cases for inclusion into the KB.

Importing the Documentation

The error code manual provided a single resolution for each error code, so we were able to translate this Framemaker™ document into HTML, break up the HTML source into segments for each error code, and import the resulting HTML page segments into the KB using ServiceSoft's import facility.

The information in the troubleshooting manual was more challenging to convert to the KB because the mapping between point-of-failure elements in the document and KB was not one-to-one. The document contained redundant information that needed to be consolidated in the KB for easier knowledge maintenance, and information important to adequately structure the knowledge base, such as proper diagnostic questions, were often missing from the documentation.

We converted the Framemaker source for the troubleshooting manual into HTML, segmented the HTML into individual cases, then ran a customized Perl script [Wall & Schwartz 1996] to map the HTML segments into KBML (Knowledge Base Markup Language), a simple SGML (Standard Generalized Markup Language) format that can be imported directly into the ServiceSoft tool. The Perl script allowed us to control the mapping between document elements and objects in the KB. In addition, by importing the new cases into a "draft" area we could integrate the new cases into the KB without losing the links between existing objects.

In January of this year, we made TS/E available internally to our capability test and environment support groups. While feedback was generally positive, our users complained that when they knew the product where the problem originated, they could not adequately use this information to narrow the search for relevant cases. We modified the system to use a user's product selection to weight entire sets of cases according to whether they were likely to contain relevant troubleshooting information. If the user didn't select a product, cases were weighted solely according to the user's query. The resulting solution search seems to be more effective at supporting their troubleshooting process.

Knowledge Acquisition

If a user cannot find a resolution for a given problem in the TS/E KB, but is able to resolve the problem using other resources, s/he can return to TS/E and complete a simple case entry form to add the new knowledge to the system. The new troubleshooting case is forwarded to the KB administrator by the TS/E Knowledge Update facility and logged to a file in KBML format. The administrator reviews the case for accuracy, consistency, and completeness before importing the KBML file into the knowledge base. This knowledge acquisition mechanism allows the TS/E system to accumulate expertise from software experts and testers prior to a software release, as well as system administrators and other users at live installations after a release. Once acquired, new knowledge is immediately available to all users on their desktop via the Web.

Conclusion

The Web is a promising medium for delivering the collected knowledge of subject-matter experts throughout an organization and redistributing it in a form more easily used by novices and is an improvement over paper documentation in many cases. Our troubleshooting expert system is always up-to-date, provides a consistent organization for the knowledge, and provides correct recommendations for difficult troubleshooting problems.

References

[Hamscher et al. 1992] Hamscher, W., Console, L., and de Kleer, J. (1992) *Readings in Model-based Diagnosis*. San Mateo, CA: Morgan Kaufmann.

[Verela et al. 1995] Varela et al. (December, 1995) DB: Browsing Object-oriented Databases Over The Web. The Fourth International World-Wide Web Conference. Boston, Massachusetts.

[Wall & Schwartz 1996] Larry Wall and Randal L. Schwartz (1996) *Programming perl*. Boston, MA: O'Reilly & Associates.

Web Advisor is a registered trademark of ServiceSoft Corporation.
Framemaker is a registered trademark of Adobe Corporation.

Creating a Workshop on the World Wide Web: The Experiences of Internet 101

Carolyn Gale, John R. Bourne, Benjamin Thompson

Center for Innovation in Engineering Education
Vanderbilt University School of Engineering
Nashville, Tennessee, USA

carolyn@dorm.net , john.bourne@vanderbilt.edu , ben@ljl.com

Abstract: This document traces the development and maintenance of an introductory workshop delivered over the Internet. Topics included using the Internet, Internet software, and building simple web sites. Results from July 1997 show nearly 700 unique visitors per week, from 60 different countries. Experiences from this workshop have formed the basis of more advanced ones on building web-based courses.

Introduction

The Vanderbilt University Center for Innovation in Engineering Education (CIEE) investigates and implements new approaches to cost-effective engineering education. We create Internet-delivered courses, experiment with tools to support on-line course authors, evaluate results, and disseminate information.

Our research involves the use of Asynchronous Learning Networks (ALN), a program sponsored by the Alfred P. Sloan Foundation [Mayadas 1997]. An ALN allows a student to learn anywhere and anytime, with information and instruction delivered primarily through the Internet and its applications.

By the summer of 1996, the CIEE had created three complete on-line courses covering basic engineering and management topics [Gale 1996]. At that point we were trying to determine if there was interest for others to learn how to build on-line courses. A grant was secured from the Sloan Foundation to develop a prototype for a future ALN Workshop on building Web-based courses. We report here our observations on building and maintaining a web-based workshop, called the ALN Workshop on Internet Basics (informally known as Internet 101).

The workshop web site is located at http://jrbnt.vuse.vanderbilt.edu/workshops/ .

Creation and Implementation

We started by assessing the current situation: limited time and resources. To minimize development time and costs, we created the hypertext documents and managed the web site using Microsoft FrontPageä , a web authoring and development tool. The FrontPage Explorer allows one to view and organize a web site, while the WYSIWYG (What You See Is What You Get) authoring capabilities of the FrontPage Editor facilitates web page development. Since

FrontPage commands looked quite similar to the Microsoft Office suite that we use, it did not take us long to learn how to use this program.

We also had no professional graphics artist; therefore, we used Paint Shop Pro, a shareware program which was useful for general artwork and screen captures of programs for tutorials.

After settling the "What tools do we use?" issue, we turned toward our prospective audience, which was relatively unknown. When we previously created web-based courses, our target audience was known: undergraduate and graduate engineering students taking required courses [Gale 1996]. Users who would view this workshop would do so on their own accord, and would certainly have a wide variety of computer knowledge and skills. We decided to form two general user groups: those who knew little about the Internet, and those who knew a fair amount.

For those who knew little about the Internet, we thought of designing a module that discussed the history of the Internet, related terminology, and security issues. For users who had difficulty using e-mail, ftp, or web browsing software, we believed that tutorials on these programs would be useful.

For those who were familiar with the above topics, we had an idea to create a module that discussed search methods on the World Wide Web, and to possibly modify a "build your own web page" laboratory from an introductory engineering class [Gale 1996].

From those general ideas, we created a rough outline that has turned into the current format:

- Part 1. Introduction / Registration / Download Needed Software
- Part 2. All About the Internet
  - So what's this Internet?
  - What else you can do online
  - Internet Security
- Part 3. Internet Software Tutorials
  - Netscape (Web Browser, Mail, Newsreader)
  - Internet Explorer (Browser, Internet Mail, Internet News)
  - Eudora (versions 1.5.4 and 3.0)
  - WinZip
  - File Transfer Protocol (FTP)
    - WS_FTP (Windows)
    - Fetch
    - Command-line

- Part 4. Finding Information on the Internet
- Part 5. Building a Web Site
    - The Barebones of HTML
    - Web Editors: What's Best for You?
    - Designing Your Web Site: Guidelines
    - Web Site Resources

Design Issues

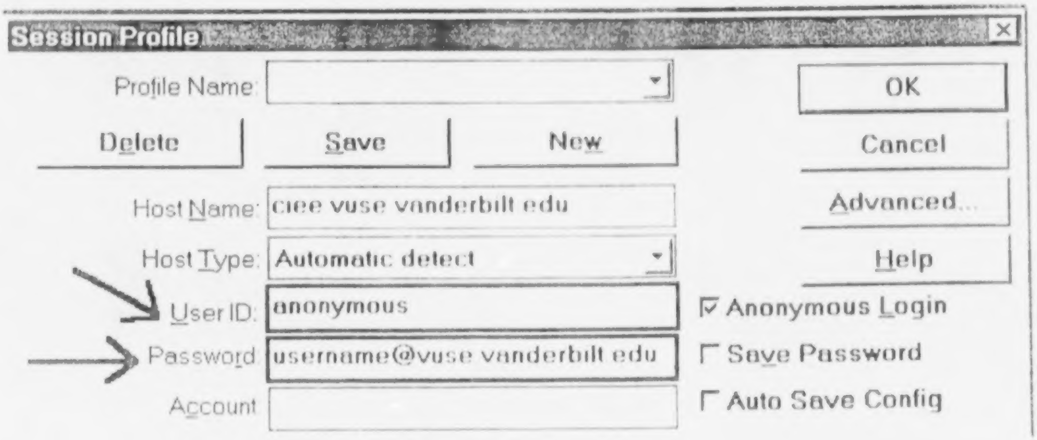
After considering the various parts to be created, we thought about the design of the web site. At first, the two students who created the workshop simply started coding in HTML, without thought to look and feel. These students modified the WS_FTP and HTML tutorials, and completed the remaining tutorials and background information in two weeks.

After being tested by co-workers and friends, the workshop was placed on a CIEE server, and announced on the ALN community mailing list in October 1996; the site was not submitted to any search engine. However, in the coming months, e-mail feedback from worldwide users forced us to consider several issues.

- **Practice what you preach.** For example, under "Designing Your Web Site", we discussed the importance of a consistent look and feel for a site; however, our site certainly didn't look that way. No one had bothered to determine the site's fonts, colors, and backgrounds. The result was that some pages had dark textured backgrounds and small fonts, others had plain white backgrounds with large ones. We finally agreed to use the latter format, and spent a fair amount of time rewriting all web pages to conform.
- **Maximize readability and minimize download time.** Users didn't want to read long pages of confusing and uninteresting text. Nor did they want to wait all day for a graphic to load.
    - First, we shortened the content of each page to prevent continuous scrolling, and rewrote many of the articles from a beginner's perspective -- this is much harder than it seems!
    - We then reduced any large graphic to sixteen colors to minimize the download time and to allow people to view the site with most monitors. Sometimes the image looked poor when we reduced the number of colors; therefore, we would change the monitor's resolution to sixteen colors and make our screen captures.
    - Finally, we tested our site by viewing it from a computer with a 14 inch monitor with the lowest resolution, using a 28.8 modem.
- **Present the user with an "interactive" example, if possible.** Many users had mentioned that case studies or examples were the best learning methods. We modified

more of our tutorials to mirror the WS_FTP tutorial, where the user FTPs to a fictional site and downloads a file. In the process this user learns how to use several of the basic commands in the program. (See Figure 1.)

Next, you must enter your complete e-mail address after you click the **Anonymous Login** box.



Host Name: ciee vuse vanderbilt edu
Host Type: Automatic detect
User ID: anonymous
Password: username@vuse vanderbilt edu
☑ Anonymous Login

You've completed the session profile. Click **OK** to connect to the remote host.

Figure 1: Portion of the WS_FTP tutorial, where one is "stepped through" a sample FTP session. A user clicking anywhere other than OK will stay right on this page.

Results

After the first version had been released in October 1996 and initial feedback came in, no one in the CIEE or the ALN community really paid a great deal of attention to the workshop.

In early 1997, we obtained a copy of Hit List, a program by Marketwave, Inc., that analyzes web site log files. Using the Internet 101 log files, we tabulated some interesting statistics, which are summarized in tables 1-3.

Total Number of Requests	12, 078
Total Number of Visits	2,311
Total Number of Visitors	1,712

BEST COPY AVAILABLE

Table 1: General results from Monday, January 13, 1997 to Thursday, February 27, 1997 (inclusive)

United States	Canada	France	Australia	Belgium
United Kingdom	Finland	Singapore	Sweden	Netherlands
Iceland	Brazil	New Zealand	Italy	Israel
Hong Kong	Austria	Norway	Switzerland	Poland
Nicaragua	Japan	Lithuania	Spain	Ecuador
Denmark	Ireland	Malaysia	Bermuda	Estonia
Germany	Luxembourg	Mexico	Portugal	Croatia

Table 2: Number of countries that visited the site: in order of most visits, read across.

Group Name	Total Requests
Internet Basics: Eudora Tutorial Main Page	1030
Internet 101	614
Internet Software Tutorials	531
Configuring Eudora	482
The Internet in a Nutshell	439
Checking Email with Eudora	421
Netscape Tutorial Home	380

Eudora Nicknames	370
Sending a Message with Eudora	370
Internet Explorer Tutorial	335

Table 3: Most popular pages from Monday January 13, 1997 to Thursday, February 20, 1997

Note that these results came with little to no advertising on our part! Using the AltaVista search engine, we also determined that 17 sites worldwide had linked the workshop, and specifically to the Eudora tutorial. Why was this?

Feedback

In order to understand why there was apparent interest in portions of the site, we developed a series of feedback forms using FrontPage. There was a general workshop feedback form, registration form, and Internet Software Tutorials form. We will focus on the Tutorial feedback form. Some of the questions asked on this form were:

- What tutorial did you use?
- Was it easy to use (Was it useful?)
- Suggestions for this tutorial, suggestions for other tutorials
- Where did you find this site?

The majority of feedback came from the Eudora tutorial. Here is a sample of what people said about the Eudora tutorial:

- Well done! Inside of 10 minutes, I had the basics. I appreciate your work very much.
- Felt it was a bit basic. I assume most people would come to changing their mail programme once they had been on the web a while and may like a bit more detail.
- Very easy to use. I did not know how to save my password and found out in about 1 minute. Thanks
- Yes since I'm very new with computers I found this very helpful.
- Easy to understand, clear and concise
- Wonderful…finally a place that explains a little about Eudora!!
- I have been a Eudora user for several months, and I picked up a couple of shortcuts looking at this. Thanks.

- Great tutorials! The Eudora tutorial was simple enough not to scare the newbies, and broad enough to get them up to speed with a minimum of fuss. The illustrations were a bit slow to load here in middle-of-nowhere rural Idaho, but they were exactly appropriate.
- I'd say I already knew about 80% by picking it up trial and error. But this was great. It was easy to follow, makes me feel much more confident that I actually understand and that extra 20% is sure going to be nice!

Suggestions for improvement to the Eudora tutorial:

- How to mail multiple recipients at the same time would have been helpful!
- Explain how to install the program after downloading
- More detail, thanks!!
- Information about attachments
- I have a Mac, would have liked specifics to it.
- Could be more in-depth

Table 4 lists some of the 96 occupations of those who used Internet 101. Note that users from all walks of life are listed there.

School Principal	Entertainer	Programmer Analyst	Banking	Nurse
Records Manager	Housewife	Church Administrator	Student	Retired
Bicycle Tour Director	Attorney	Benefits Coordinator	T'ai Chi Teacher	Systems Analyst
Accountant	Journalist	Office Furniture Sales	Financial Consultant	Software Trainer
Manufacturer	Pilot	GM Auto Worker	Physician	Marketer
Canadian Armed Forces chef	Writer	Tax analyst	Electrician	Waitress

Table 4: Occupations of Internet 101 users.

We have visited the Eudora web site and found links to sites with Eudora resources. We emailed the site managers and had our site linked there. Also, we posted a message on relevant Eudora newsgroups and asked for feedback. This started to attract more visitors to our site.

In the next month, statistics increased dramatically (see Table 5). This was after submitting the Eudora Tutorial site to Yahoo, and Internet 101 to AltaVista. As of July 16, 1997, we had users from 60 countries on 6 continents. The Eudora tutorial has nearly reached 5,000 visitors, and Internet 101 in general has reached nearly 2,000.

Total Number of Requests	52,328
Total Number of Visits	9,344
Total Number of Visitors	7,042

Table 5: General results from Monday, January 13, 1997 to Wednesday, July 16, 1997 (inclusive)

Modification

From this feedback, we added new tutorials and revised incorrect or unclear information. Information comes to us from one or two feedback entries per day and the occasional email. Information on the Netscape and Internet Explorer modules were upgraded to the latest version, and sections on creating signatures, attachments, and mailboxes were added to the Eudora tutorial. We keep the tutorials section updated as often as possible, and have added command-line FTP, Fetch, and WinZip to the tutorials list.

We haven't worked with other workshop sections in great detail, since we want to focus on the sections that are used most often.

Future Plans

Many users have written requesting copies of the Internet 101 material, or permission to add our URL to their site. At the end of July, 1997, we started to sell the complete workshop for users who wished to customize the material. Early sales have been promising – our first sale came shortly after the order form was placed on-line!

Other plans are to expand the number of available tutorials, and to provide more in-depth information for those who desire it.

This prototype workshop was used as a readiness module for an expanded workshop on building ALN courses. This new workshop was offered as part of the Third International Conference on Asynchronous Learning Networks from August – October, 1997.

Conclusion

The workshops that were created in the summer of 1996 have become more popular than we had expected. Starting with a rough outline, we created a series of modules to support users who want to learn the fundamentals of the Internet and software used on it. We also included information and examples of building basic web sites.

Analyzing results in early 1997 showed that many people had found the workshop and were linking the site to their own. Placing feedback forms in several areas of the workshop helped fine-tune our materials (in many cases, add more information) to satisfy our users. We have used this workshop as a basis for advanced workshops on building web based courses.

References

[Gale 1996] Gale, Carolyn M., Bourne, John R., Dawant, Martine M., Quinn, Doris (1996). Implementing Engineering Education on the WWW: 3 Case Studies. WebNet: World Conference of the Web Society, 1996, Association for the Advancement of Computing in Education, Charlottesville, VA.

[Mayadas 1997] Mayadas, Frank. Asynchronous Learning Networks: A Sloan Foundation Perspective (1997). Nashville, TN: Journal of Asynchronous Learning Networks, 1, 1-17. [http://www.aln.org/journal/issue1/mayadas.htm]

Acknowledgements

Susan Hall, a senior at Brigham Young University, for creating the Fetch tutorial, and Kevin Tangney, a senior at Vanderbilt University, for creating the WinZip tutorial.

Sarah Stockwell, at Qualcomm, Inc.'s Eudora division, for supplying technical advice for the Eudora tutorial. Jeff Ross also provided contacts to other Eudora-related sites.

Martine Dawant, Research Instructor of Computer Engineering at Vanderbilt, for providing initial workshop feedback.

Members of the Spring 1996 Informatics class at Vanderbilt University (Geoffrey Reynolds, Wesley Williams, Larry Kwok) for creating the prototype computer literacy course. Portions of the project would be the basis for the software tutorials.

Special thanks go to all who took the time to complete the feedback forms — their advice was invaluable.

Development and Integration of WWW-based Services in an Existing University Environment

John Garofalakis, E-mail: garofala@cti.gr

Panagiotis Kappos, E-mail: kappos@cti.gr

Athanasios Tsakalidis, E-mail: tsak@cti.gr

John Tsaknakis, E-mail: tsaknaki@ceid.upatras.gr

Giannis Tzimas, E-mail: tzimas@cti.gr

Vassilios Vassiliadis, E-mail: BB@cti.gr

Department of Computer Engineering and Informatics,
University of Patras and Computer Technology Institute, Greece

Abstract: This paper presents the experience and the problems solved by our implementation group, in the process of developing and integrating advanced WWW-based services in the environment of a moderately sized University of Greece, namely University of Patras, which by now offers only basic network services (e-mail, ftp). In the following we present a short overview of the services developed, the overall system architecture, and the critical aspects of introducing the new services to the users.

1. Introduction

The introduction of advanced network services into a university environment is today a basic need, the satisfaction of which enables the leverage of the campus administrative operations, the collaboration between different scientific groups within the university providing new means of communication and introduces the use of new teaching methodologies via the network [December and Randall, 1996]. However, it is not an easy task since it has to overcome the traditional ways of administration, information sharing, and teaching. Moreover, it needs an effective user-oriented implementation and support mechanism in order to assure its widest acceptance and use by the academic community [Reinhold 1996]. At the time being, the University of Patras only supports basic network services such as e-mail and ftp and a few WWW servers developed within some of the University's departments, that partially support the whole campus needs, whereas services like on-line and off-line tele-training and videoconferencing only exist in an experimental level in some of the laboratories.

The basic aim of our project is to provide a set of advanced network services in the campus of Patras. The key point in this effort is to provide the whole set of services under a uniform platform, that is to integrate the services into a system using WWW technology. Beyond the basic services (e-mail, ftp etc.) that are going to be implemented within this project the final system will integrate the following set of advanced services:

- A WWW-based information service.
- Intranet services to support the administrative operations within the campus.
- Distance learning by means of on-line and on-line tele-traing via the Web.
- Teleworking facilities.
- Videoconferencing facilities.
- Applications supporting collaborative work.

The whole system will be realized through the use of the University network which will be based on the TCP/IP protocol technologies enhanced by the 100Mbit speed, obtained by the fiber optic lines (FDDI) used to connect the University backbone. Two ATM switches will be exploited to connect the high demand real-time applications such as video conferencing.

2. Services Provided by the System

The exploitation of WWW technologies will be the base upon which the final system will be developed. The central web server of the University will provide a wide range of services such as:

- Information about the institution as well as general information such as announcements, festivals and other social activities in the form of multimedia rich documents.
- Links to all other departmental web servers in order to reflect the current status of all the departments.
- A powerful search engine aiming to provide an easy-to-use interface for locating information based on keywords.
- Collaboration tools (such as customized USENET News or bulletin board Software) for information sharing
- Mail services with multimedia capabilities (voice mail)
- A uniform and sophisticated way of updating or inserting information, in order to give all users (professors, post-graduate or under-graduate students) the potential of information publishing

The whole system will be developed based on third party public domain or freeware software (APACHE web server, Harvest and HtDig etc.). By using the latest programming techniques such as JAVA, JavaScript, ActiveX, VRML and WYSIWYG HTML editing, an interactive interface will be built which will enable the use of multimedia in all laboratories of the University [Stone 1994] [Chee 1996].

Another service that is going to be developed within the project life-cycle will be the implementation of several Intranets within the campus, aiming at the reduction of paper use in the administrative procedures in the University. All the traditional paper-only distributed sheets or books will be stored electronically. Using text search and efficient retrieval techniques all documents will be delivered on-demand to named groups of authorized persons without any bureaucracy [Bernard 1996].

A significant aspect that arises in this case is the protection by intruders from outside the University network or from unauthorized users. The encryption provided by the SSL 3.0 protocol will be exploited to transmit information securely. Flexible user authentication controls, read/write access to individual files or directories using user name and password, domain name, host name, client-side certifications or named groups will be exploited.

Videoconferencing & Tele-training will be included in the set of the advanced services provided by the final system. These services enable real-time conferencing interactions over the Internet and Intranet. Conference sessions will allow the University to increase the effectiveness of workgroup, departmental, and cross-functional communication by letting users interact on the same documents, sketching on collaborative whiteboard, exchanging data files, and talking in real time with colleagues in or outside the University [Bouras 1996a].

Customized software will be constructed to enable all university users to participate in Videoconferencing (on-line, off-line) sessions [Basiroglou 1991]. Off-line Videoconferencing will include pre-recorded material such as a tutorial of classroom course. The course will be embellished with pictures and/or video files to give the attendees the closest possible impression conveyed inside the classroom the actual course was given. On-line conferencing refers to the real-time transmission of audio and motion images to multiple recipients. IP Multicasting technology, in conjunction with the latest H.323 and RTP standards, will be exploited to provide timely crucial data [Bouras 1995] [Bouras 1996b].

A Realaudio server will be installed to host all the voice announcements, extracts of important conference speeches and music or other voice material. This server will provide easy voice information access to not only low-speed dial up users but to all other directly connected nodes. The compression and streaming will save precious bandwidth for other applications.

Finally, remote users will be able to access the University network facilities by remote access services. Two kinds of remote access have been defined. The first kind is using the conventional digital phone lines media of communication (33.6KB or 57.6KB modems). Users of this kind will be satisfied at reading multimedia mails, net-surfing the world-wide web sites, Internet chatting, accessing bulletin boards or transferring files. The second, is using ISDN. ISDN access will be supplied to users, who need high speed access to multimedia services with real-time response such as on-line Videoconferencing services.

3. System Architecture

The WWW services will use the Client - Server model so as to take advantage of its ability to distribute data and processing chores across the campus network. The main parts of the application and services run on centralized servers, and any user may have control using special client software designed for this purpose [Nicolaou 1990]. Thus, a number of servers have to be implemented for the provision of the vast volume of information for every department of the University of Patras. Storing and distributing this information using only one server, is not a good solution for a number of reasons:

- the ever growing volume of information originating from the large number of departments of the university, will certainly pose storage problems
- the expected large number of visitors in the web pages of the university server is expected to slow down considerably its network performance
- possible malfunction of the central server will result in the total suspension of every WWW service

Having in mind the above parameters the physical architecture of [fig. 1] has been chosen for the implementation of the services. A central server will store general information concerning the University (historic, geographic information) and links to other servers, which operate in every department of the campus. Similarly, the servers of each department will contain information for the department and links to laboratory WWW servers. Each laboratory will use a separate server for the publication of its research achievements, along with various technical and educational information.

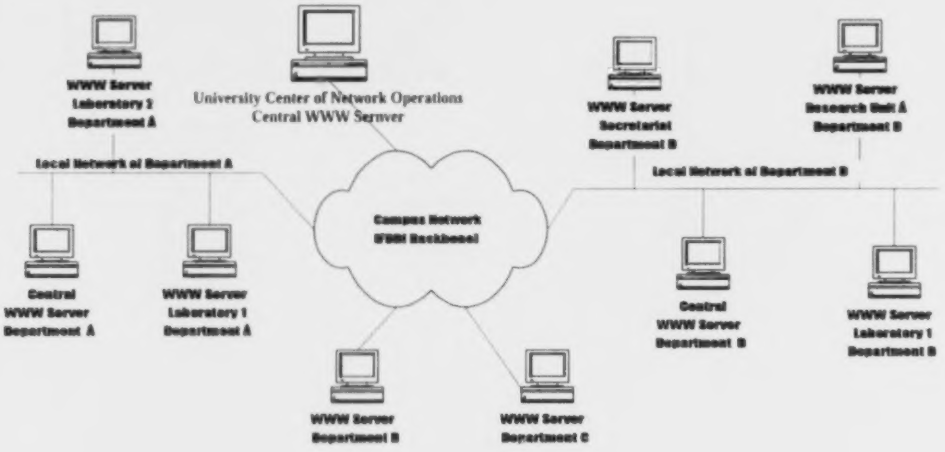

WWW Server
Laboratory 2
Department A
University Center of Network Operations
Central WWW Sernver
WWW Server
Secretariat
Department B
WWW Server
Research Unit A
Department B
Local Network of Department A
Local Network of Department B
Campus Network
FDDI Backbone
Central
WWW Server
Department A
WWW Server
Laboratory 1
Department A
Central
WWW Server
Department B
WWW Server
Laboratory 1
Department B
WWW Server
Department B
WWW Server
Department C

Figure1: Representation of WWW Servers in campus network

The Client - Server architecture is distributed and results in flexible network structures. A main advantage is the capability offered to each laboratory or research team to control all the needed information independently and in a very efficient way. Each department or laboratory will have total control of the provided services causing the traffic load of the campus network to be equally distributed, increasing thus the total network performance.

It should be noted that a laboratory server is not a dedicated WWW machine. Due to the relevant low traffic expected for each laboratory server, standard computer equipment will be used for this purpose. Another alternative is the virtual host implementation, where multiple laboratory servers will be hosted in a single machine. Department servers may also be temporarily hosted in the central university server. The client server architecture does not require special infrastructure or investment by any department or laboratory for its implementation.

The WWW clients are installed in workstations (personal computers or Unix machines) and every user can access both local and remote WWW servers (of other departments or universities).

The implementation of the WWW services requires the use of special transport and control protocols for the handling of information. TCP/IP will be used as the standard communication protocol between the clients and the servers along the network. Initially, the services will be developed and tested in a laboratory LAN. The open architecture used in both the communication protocol and the services ensures the proper operation in the

BEST COPY AVAILABLE

university's WAN, which is using FDDI, ISDN and ATM technology [Wolfinger and Moran 1991] [Shepherd 1992] [Grudin 1996].

Each WWW server uses the HTTP (HyperText Transport Protocol) for the transfer of data (text, images and sound) to/from the network. The use of HTTP allows the communication between the WWW server and the client (a WWW browser) via a socket connection established by the TCP/IP protocol [Newcomb 1991].

As far as data security is concerned, the use of special transport protocols such as SSL (Secure Sockets Layer) or/and S-HTTP (Secure HTTP) ensures the transfer of confidential information through secure channels. Such a need in the University of Patras rarely arises, but even then other methods such as authentication based on the source network address and passwords meet, to some point, the needs for security [Garzotto 1993].

CGI (Common Gateway Interface) is the most common way of communication between Web applications and Databases, creation of search engines and presentation of web pages. It will be used for the implementation of services, which require a more powerful implementation tool than HTML [Gebhardt 1995].

The mail services will use a variety of protocols including SMTP (Simple Mail Transport Protocol), MIME (Multimedia Interface Mail Extensions) and POP3 (Post Office Protocol). These protocols will be used for the transfer of messages via e-mail or distribution mailing lists. The inclusion of MIME enables the transfer of not only text but of multimedia messages as well [Costa Carmo 1992] [Bulterman and Liere 1991].

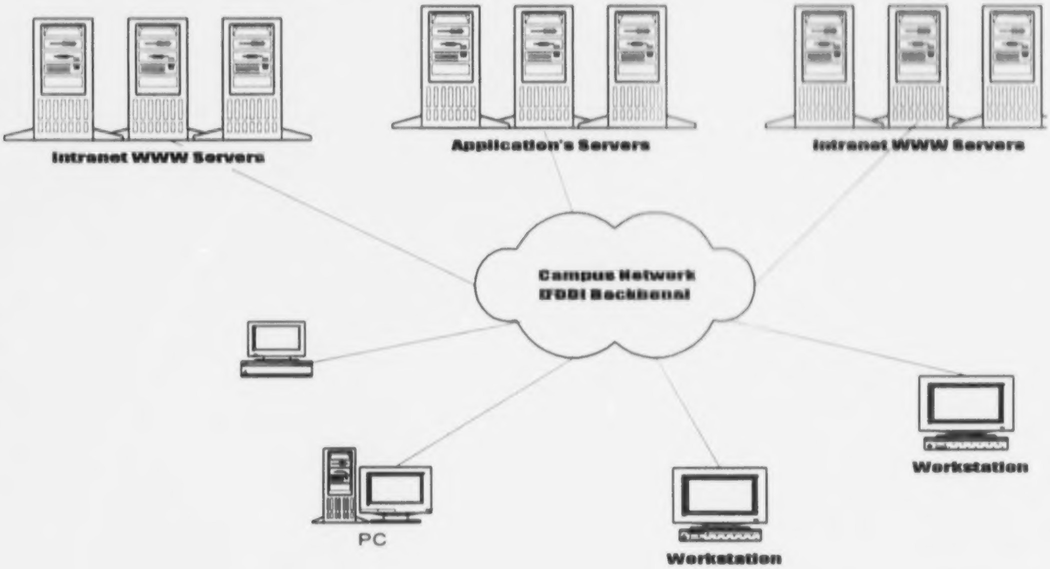


Figure 2: Intranet Architecture

Finally, the Intranet architecture will use the client - server model and the same protocols as well. [Fig. 2] represents the Intranet infrastructure at the University of Patras.

4. Implementation and Introduction of the Services to the Users

One of the most critical stages of the whole project is the introduction of the services to the users. International experience has shown that the gradual and easy introduction of the system to the users as well as its interactivity and functionality are some of the major factors that will determine its acceptance. Moreover, having in mind that the final system will be used for the educational procedures within the campus, several pedagogical aspects must be taken into account.

Based on the above considerations the project team is going to consume a great deal of efforts towards the following directions:

- The administration and support of all network services will be integrated in the University Center of Network Operations.

- Integration of the whole set of services under a uniform platform using a friendly and easy to use user interface to support the interaction with the users.
- A special team of pedagogues will design the entire human-machine interaction, especially in the case of distance learning.
- A common methodology will be developed for the implementation of similar services in every department of the campus.
- There will be on-line help available, as well as a special team that will support the users in the case of any technical or non-technical problems.
- A series of seminars will be held for the introduction of the services to the users

The final purpose is to develop an interactionally rich system to support efficient and effective user functionalities, taking advantage of the new WWW multimedia infrastructures currently available based on a friendly and easy to use user interface.

5. Conclusions

We will develop, at the University of Patras, a set of advanced services to facilitate academic and research activities. Internet applications are worldwide used to support all relevant activities. Web pages developed will include a number of information and search engines will be used to provide access to these pages. Although, Intranets are not widely used in academic environments, but recently a tension has aroused in developing such services (especially in the USA) to support intercommunication between the different departments of the University of Patras. The services to be developed will serve as a guideline for all relevant applications to be developed in the future in Greece

Our work showed once more that the introduction of new network services in an existing environment, even if this is a University, is not mainly a technological problem. The development of the services, their integration and introduction to the users has to follow a well-defined, user oriented implementation plan.

The final output will be a pool of advanced services focused on the user needs for effective information retrieval and spreading, and the use of alternative education tools

References

[Bernard 1996] Bernard, R. (1996). *The corporate intranet.* Wiley Computer Publishing

[Reinhold 1996] Reinhold, V. N. (1996). *Spinning the Web: How to provide Information on the Internet*

[December and Randall, 1996] December, J., & Randall, N. (1996). *The World Wide Web Unleased.* Sams Publishing

[Garzotto 1993] Garzotto et al., (1993). Navigation Patterns in Hypermedia Data Bases . *Proceedings of 26th Hawaii International Conference on System Sciences, 1996*

[Newcomb 1991] Newcomb et al., (1991). The HyTime Hypermedia/Time-based Document Structuring Language. *Communications of the ACM*

[Costa Carmo 1992] Costa Carmo, L.F.R da, Saqui-Sannes, P. de, Cuartiat, J.P. (1992). Basic synchronization concepts in multimedia systems . *in Proc. 3rd International Workshop on Network and OS Support for Digital Audio and Video*. 514-518.

[Bulterman and Liere 1991] Bulterman, D.C.A, Liere, R.van. (1991). Multimedia Synchronization and UNIX. *In Proc. 2nd International Workshop on Network and OS Support for Digital Audio and Video*, 109-119.

[Nicolaou 1990] Nicolaou, N. (1990). An Architecture for real-time communication systems . *IEEE Selected Areas In Communications*, vol8 ,no3

[Wolfinger and Moran 1991] Wolfinger, & Moran, M. (1991). A continuous media data transport service and protocol for real-time communication in high speed networks . *In Proc. 2nd International Workshop on Network and OS Support for Digital Audio and Video*, 171-182.

[Shepherd 1992] Shepherd, D., Hutchinson, F., Garcia, G., Coulson. (1992). Protocol support for multimedia applications. *Computer Commun.*, July 1992, vol.15, 359-366.

[Stone 1994] Stone, J. (1994). An Empirical Study of Delay Jitter Management Policies. *Multimedia Systems, 1995, vol.2, Number 6*, 267-279

[Stallings 1988] Stallings, W. (1988). *Data and Computer Communications.* Second Edition, MacMillan Publishing Company

[Basiroglou 1991] Basiroglou, P., Bouras, C., Garofalakis, J., Papoutsopoulos, G., Spirakis, P. (1991). RECPHONE: A new environment for medical remote expert consultation. *EuroPACS'91, 9th International Meeting*, Berlin, July 2-3

[Bouras 1996a] Bouras, C., Fotakis, D., Kapoulas, V., Kontogiannis, S., Kyriakou, K., Lampsas, P., Spirakis, P., Tatakis, A. (1996). An Interactive Cooperative Teleworking Environment- Ô çëå Ü èåé á *ED-TELECOM 96, June 17-22 1996*, Boston, USA.

[Gebhardt 1995] Gebhardt, M., Jacobs, S., Kethers, S., Rzasa, W. (1995). Filling HTML Forms Simultaneously: CoWeb - Architecture and Functionality. *Computer Networks and ISDN Systems*, Volume 28, issues 7 11, 1385

[Grudin 1996] Grudin, J. (1996). CSCW: History and Focus. *Information and Computer Science Department*, University of California, Irvine

[Bouras 1995] Bouras, C., Fotakis, D., Kapoulas, V., Lampsas, P., Papoutsopoulos, G., Spirakis, P., Tatakis, A. (1995). HIPPOCRATES: A multimedia tool for distance education. *ED-MEDIA 95, June 17-21 1995*, Gratz, Austria

[Bouras 1996b] Bouras, C., Kapoulas, V., Kastis, N., Spirakis, P. (1996). TRENDS: Training Educators through Networks and Distributed Systems. *EDEN 1996, 8-10 July 1996*, France

[Chee 1996] Chee, Y.S. (1996). MIND BRIDGES: A Distributed Multimedia Learning Environment to Support Collaborative Knowledge Construction. *International Journal of Educational Telecommunications*, 2(2/3), 137-153

Adaptive and Agile Interactive Learning Environments on the WWW

Sylvain Giroux
Cognicase
1080 Beaver Hall,
Montreal, P.Q., Canada,
e-mail: girs@cognicase.ca

Richard Hotte and Kim Dao
LICEF Research Center,
Télé-Université,
1001, Sherbrooke east,
Montreal, P.Q.,
Canada H2L 4L5
tel.: (514) 522-3540 ext. 4276
fax: (514) 522-3608
e-mail: {rhotte, kdao}@teluq.uquebec.ca

Abstract Continuous education helps people to cope with an ever changing labor market, while distance education reaches them where they are, keeping them at work. We designed a framework used to producing learning environments (LEs) on the WWW. The resulting LEs are germane to fractals. First, we liken changes in scale to levels in LEs. Each level expresses a given viewpoint on knowledge. Second, self-similarity establishes a classification from which to derive a grammar. Third, texts and activities are highly fragmented. Fourth, the interfaces rely on the fractal structure to provide for "spatial" landmarks. The LEs are *adaptive* with respect to learners' objectives, background and cognitive style and are *agile* with respect to their design, implementation and maintenance. The fractal design and the underlying grammar set up the formal grounds required to code procedures that generate LEs, extend them, manage updates, and maintain the site.

1 Introduction

Nowadays people cannot afford to stop working in order to update their education. At the same time, the fast evolution of technology and an ever changing labor market impel them to keep their knowledge up-to-date. Distance education offers a satisfying solution to issues related to continuous education by reaching peoples were they are, keeping them at work and helping them to cope with a changing world. Since the mid 70s, Tele-university has produced pedagogical material for learning at a distance including texts, videos, exercises, exams, teleconferences, assistance, and other options. All these elements define what we will call a learning environment (LE). But producing high-quality pedagogical LEs is a long and costly process. This paper presents a framework that improves productivity and quality at a reduced cost for both the designers and the learners.

On the design side, the framework provides solutions to the reuse and maintenance of pedagogical material. It defines procedures for updating the pedagogical material in order to take into account the evolution of knowledge, and also gives a methodology and the means to tailor a course or part of a course for customized training. On the learner side, an LE has to interact with a variety of learning styles. Our framework enables one to produce LEs where a learner is free to select the part of knowledge he learns, free to choose the way she explores the pedagogical material, and free to choose the place and the time she studies.

Hence, the LEs generated by the framework are adaptive with respect to a learner's needs, background and cognitive style while they are also agile with respect to their design, implementation and maintenance. We used this framework to design and implement two hypermedia LEs on the WWW:

INF6550 LE, called *Methods and Tools for Problem Solving*, is intended for undergraduate students having a background in business;

TEC6200 LE, called *New Information Technologies and Cognitive Development*, is intended for graduate learners having a background in education.

Hypermedia is the cornerstone of our LEs' agility and adaptiveness. The WWW provides means for real-time modifications of their hypermedia structure.

In this paper, we detail this framework by explaining how we designed these LEs and how we exploit their adaptiveness and agility. First, we explain the hypotheses underlying our design, especially since they provide the basic assumptions that pointed out the supporting tools we implemented. Second, we sketch the fractal structure that is at the heart of the systems. Third, we show how to take advantage of this structure from the standpoint of the learner to adapt the LE to his or her needs and cognitive style. Fourth, we switch to the point of view of design and implementation. We show how the hypermedia structure can provide agility to production. Fifth, we discuss the benefits and drawbacks we observed from in situ utilization of the LEs. Finally, we sketch future works and draw concluding remarks.

2 Towards Constructivist TeleLearning Environments

Tele-university is devoted to distance education. Since its creation, courses have incorporated a variety of media. More recently, telelearning environments broke learner isolation and provided means for distributed learning. Besides giving access to knowledge, these environments help learners to manage their learning process, or to communicate with their peers. Collaboration then takes the form of discussion, virtual teamwork, or asynchronous assistance by the tutor. The information highway reveals as the latest challenge to Tele-University: how to produce fully computer-mediated learning environments. We claim that the INF6550 and TEC 6200 LEs lay down the basis of the answer.

Learning is unquestionably a complex phenomenon. Many elements and processes are interwoven. To be efficient, an LE must integrate them smoothly; even we could say that it should provide for some sort of symbiosis. To help us to get a clearer view, let's put them in a three-orthogonal axis:

- *static elements*: they consist mainly of texts, pictures, sounds and videos that explain theory, give examples, describe exercises and point out tools which the learner can use;
- *dynamic processes*: while learning, a person undertakes many actions which reflect the learning process by itself;
- *assistance*: when stuck on a problem, a learner searches for help; at Tele-university a telelearner gets support from a computer-mediated support system that achieves asynchronous communication between individual learners, networks of learners, and tutors to ease truly cooperative telelearning [Pierre & Hotte, 96]; in a near future, advisor information systems will provide first-line help [Giroux et al., 96].

So the challenge was to mediate these three aspects and generate a complete Web-based LE. Even better, we aimed at taking advantage of the flexibility and interactivity the Web offers (e.g., the hypermedia structure) to give a learner complete freedom on the knowledge he chooses to learn and on the manner he learns it. The latter is often referred to as a constructivist approach to learning.

In building an LE, static elements, dynamic processes and assistance raise their own set of issues. Static elements settle the playground. How should we structure and present contents and activities in a Web-based LE? What are the (html-)pages to produce? Dynamic processes correspond to the way a learner uses static elements. How are the static elements used? How does constructivist learning express itself? How could an LE support each person's learning process idiosyncrasies? Remote assistance is crucial in distance learning. How would an LE support learners that are free to choose their progress? Which are the help resources that are relevant and appropriate?

At first sight, these questions seem unrelated and consequently one could expect to solve each one on its own. But they are deeply intertwined. For instance, the freedom a learner has is intimately linked to the content and format of the documents. Longer documents leave less room to freedom since the learner usually has to read them all, thus putting more constraints on the order knowledge is acquired. We thus ought to uncover the relations existing between static elements, dynamic processes and assistance.

Our stance over Web-based LE is summarized in the following hypotheses that link contents, learning, assistance and learners' freedom:

1. Constructivism approach focuses on the learner. Constructivism requires that the LEs give complete autonomy and freedom to a learner's thought process. Such freedom especially implies that the learner should be able to interact with the LE according to his or her own cognitive style.
2. Hypermedia can provide such freedom to the learner.
3. To ensure coherence within hypermedia learning system, it is compulsory to establish a symbiosis between contents (knowledge), form (documents and activities) and though processes (cognitive progression and assistance).
4. To build constructivist hypermedia learning system, it is possible to lean on the symbiosis between contents, form and though processes.
5. The learning process (and thought process in general) is reflected by the navigation of the learner in the static elements.
6. It is possible to develop supporting tools, especially tools that render explicit the learner's route and progress in constructivist hypermedia LE.
7. To support constructivism, tools that render explicit the learner's route and progress are required.

For any feature in the LEs, one can report to some of these hypotheses. For instance, assumption #3 is underlying the complete design of the hyperstructure. Tracing tools were implemented in light of assumptions #6 and #7.

3 Elements of Adaptive Telelearning Environments

Indeed, we are looking for telelearning environments on the WWW able to adapt to a learner's objectives, background and cognitive style. The hypermedia structure underlying the WWW provides the basis for the kind of adaptation we are interested in. In this section, we first point out the need for design principle in the realm of hypermedia. Then we describe the principles fractals that rule the structure of the LE we implemented. Once the fractal structure is set, we know both which documents to write and what is their content. The hyperlinks can then be derived based on the fractal structure. The fractal structure is also used to design the interface. Once the static part is defined, the stage is set to study the dynamic processes and tools are designed to support the learner's cognitive processes. Finally we show how the design-produced documents, i.e. WWW pages, enable a learner to adapt the LE to his or her own cognitive style just by the way he navigates through them.

As the supporting medium for LEs, the WWW has many advantages regarding agility for distance education (e.g., distribution, real-time modification and notification). On the other hand, this medium imposes severe constraints. Learners have to work using a usually small computer screen. Linearity in thought and texts rapidly becomes quite boring. Consequently, designing pedagogical material for the WWW is far from writing a textbook. In textbooks the unit of division is, roughly speaking, the section. The structure is linear. When an author writes a section, he usually assumes that the reader will have knowledge of the preceding section. Books have been written for centuries, and today there are guidelines to achieve such a process. For instance, tables of contents are part of the conventions guiding the organization of books.

The WWW is a very young media, and there are no universally accepted guidelines. On the WWW, the unit of fragmentation is the page which, we believe, should be no longer than two computer screens. The structure is hypertextual, so the author cannot predict which path will have lead the learner to the page he is writing. Thus, a page should address one micro-idea. Due to the medium, a small computer screen and relatively low-bandwidth for communication, documents must be kept short. Finally, documents ought to be self-explaining. But these guidelines are not sufficient. A WWW author needs also guidelines to provide answers to the following questions:

1. What are the pages he has to write down? What content and knowledge should each page address? ;
2. How should contents and activities be linked to obtain LE in such a way that the learning process remains open?.

The answer comes from the very nature of hypermedia: fractals!

Hypermedia provide a very powerful and flexible mean to present knowledge to the learner [Jonassen, 86]. But as the WWW reminds us every day, it is quite easy to get lost in such fragmented universes. So we sought for an organization of didactic material that could ensure coherence throughout the material. Fractals possess the qualities required to organize highly fragmented universes such as those found on the WWW:

> "A fractal is a rough or fragmented geometric shape that can be subdivided in parts, each of which is (at least approximately) a reduced-size copy of the whole. Fractals are generally self-similar and independent of scale" [Stepp, 96].

We believed (and we observed afterwards) that a structure inspired by fractals can help to ensure coherence between the various hypermedia fragments by defining a sort of spatial relationship. The spatial structure obtained gives landmarks to learners to help them avoid getting lost. The issue then is to organize pedagogical material into self-similar levels of interrelated fragments. On the one hand, we liken changes in scale to levels in LEs, each level expressing a given viewpoint on the same knowledge. On the other hand, self-similarity establishes a classification. We used this classification to derive grammars and to provides formal grounds for a methodology and for automation. Let's see how it is achieved.

An LE is made up of pages. A page content may be theoretical (description of theory), pragmatic (description of activity), or related to the LE by itself (for instance, pointers on help resources). The fractal structure of an LE focusses on pages that either describe the theory (models, examples...), or activities (exercises, homework...). We call each page a *fragment*. Even if fragments describe independent micro-ideas, they are linked to each other. Then, the links between the fragments create a network. Whatever complex the network is, some nodes, usually a few, are fundamental and the rest of the network can be interpreted as examining these nodes from a different perspective or as a finer grain view of them.

Besides the network of fragments, there is another network implied in an LE. Knowledge addressed in it could be modeled as a rich semantic network. We called it the knowledge model [Fig. 1, left]. Any fragment, theory or practice handles some portion of the knowledge model. The trick to get to a fractal structure of the LE is to find a mapping between fragments and the knowledge model. The viewpoints on a subject determine the levels, while the central nodes of the knowledge model define the main part of self-similarity [Fig. 1, right]. The other elements defining the grammar and completing self-similarity are the type of the fragments .theory or practice. and the subtype of the theoretic fragments .presentation, model, examples.. Thus, levels and self-similarity provide the guidelines needed to determine which documents to produce and define a standard way to fragment knowledge and to ascribe a topic to individual fragments. Even better, levels and self-similarity define the foundations of a grammar that indicate the documents that have to be produced.

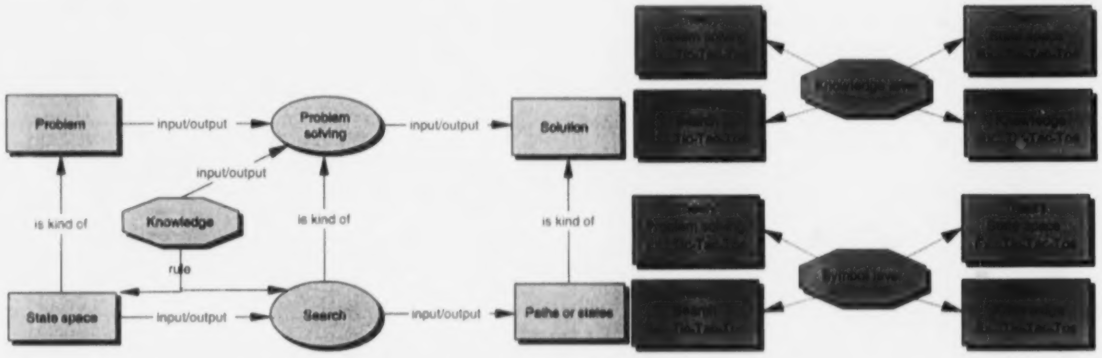


Figure 1: Left) The knowledge model of the course INF6550: to solve a problem is to perform a search in a state space; the search is directed by knowledge [Newell and Simon, 76]. Right) Self-similarity between levels is derived in part from the knowledge level.

Once fragments are written, they have to be linked. Since knowledge is at the heart of the LE, knowledge is the main criterion used to define a coherent hypertextual structure. The problem consists in identifying for each fragment the fragments that are semantically the closest. Such information is made explicit by the fractal structure, based on the knowledge model, the levels, and the pedagogical nature. Since this information is encoded in the name of the fragment, the computation of the semantically closest fragments can be done on the fly. This property enables one to distinguish in the interface two types of links according to their semantic. Links with a

BEST COPY AVAILABLE

strong semantic connotation are implemented with the help of a contextual navigator [Fig. 2]. Links with a weak relation to the discourse, as references, are implemented as usual.

For instance, in the INF6550 LE, the contextual navigator can be thought of as a hypercubic structure based on :

1. the axis of the knowledge model: problem-solving (S), knowledge (C), search (R) and state space (E);
2. the levels: objectivation, knowledge, symbol and expertise. The forefront plane correspond to the current level;
3. the pedagogical nature: theory and practice. The theoretical fragments are associated to the upper halves of the small squares, whereas the practical ones are associated to the lower halves.

Other elements as the example name is not described explicitly by the contextual navigator. The shaded part indicates the semantic of the current fragment. The contextual navigator indicates that the fragment is about the state space from a pragmatic approach (lower part is shaded) at the symbol level. Finally, some other contextual elements are made explicit through links established beforehand: knowledge models, examples list...

In order to manage its learning process, the learner must be able to locate himself among a bulk of knowledge fragments. Besides the information made available by the contextual navigator, other indicators have been incorporated in the interface[Fig. 2]. Very quickly, learners notice that the level of fragment is identified by a specific color stripe : blue for problem-solving, yellow for search, pink for knowledge and green for state space. At the upper left corner of the document, there is always an icon to indicate the level. Finally, each document has a title providing further information. Conventions govern title wording. From now on, the hyper-stage is set, and the learner can come in.

In a constructivist approach, learning processes need to be observed and taken into account [Chambreuil et al., 94]. To manage her learning, the learner must know what she learned and what rests to be learned. To give her feedback on its learning process, we implemented a trace mechanism. The trace indicates which fragments have been visited, and to what extent they are understood or completed. The trace uses a six-steps scale: introduction, planning, beginning, entry, strengthening and complement. A color is associated to each step. The progress of the learner is estimated automatically, but the learner can always indicates to the system what its real progress is.

The trace is displayed either on a fragment-per-fragment basis or on a synthetic map. In the first case, an arrow on the scale at the upper right of each fragment indicated the inferred or real degree of understanding. In the latter case, a global cognitive map gives a synthetic view of the LE's fragments, level by level. Fragments are represented by cells whose color indicates the learner's progress. The synthetic map also helps the learner to appraise how many fragments there are for each level, as well as how many are theoretical or practical.

There are many ways to navigate within an LE. We have already explained the contextual navigator which points on the fragments that are semantically the closest. The contextual navigator can be used to follow a line of thought. But there are times when the learner may want to stop investigate an idea and jump to another one. She can also navigate through the hypermedia LE using: the synthetic map, a menu at the bottom of the fragment , or specific items pointed out in the current level.

Now the learner has at hand LEs that are highly fragmented while still remaining well-structured. There are clear division between theory and practice, and the essential features in the knowledge model are highlighted. There are also contextual maps, synthetic maps, and trace recordings to give her a view on her learning process. But what about this learning process? How could she navigate through the LE according to its cognitive style? A free translation of the Myers-Briggs type indicator can provide some hints [Krebs, 85]. People may be either theoretic or practical, synthetic or analytic... For instance, the theoretical one will consult the theoretic fragments first; then he will do the exercises, while the analytical one will choose a theme and will follow it throughout all the levels. The LE is sufficiently structured, flexible enough, and will provide the right tools to enable a learner to adapt it to his own cognitive style. Finally, asynchronous transactions between tutors and students are privileged in the LE. So the learner remains free to choose the time and place most suited to its life-style.

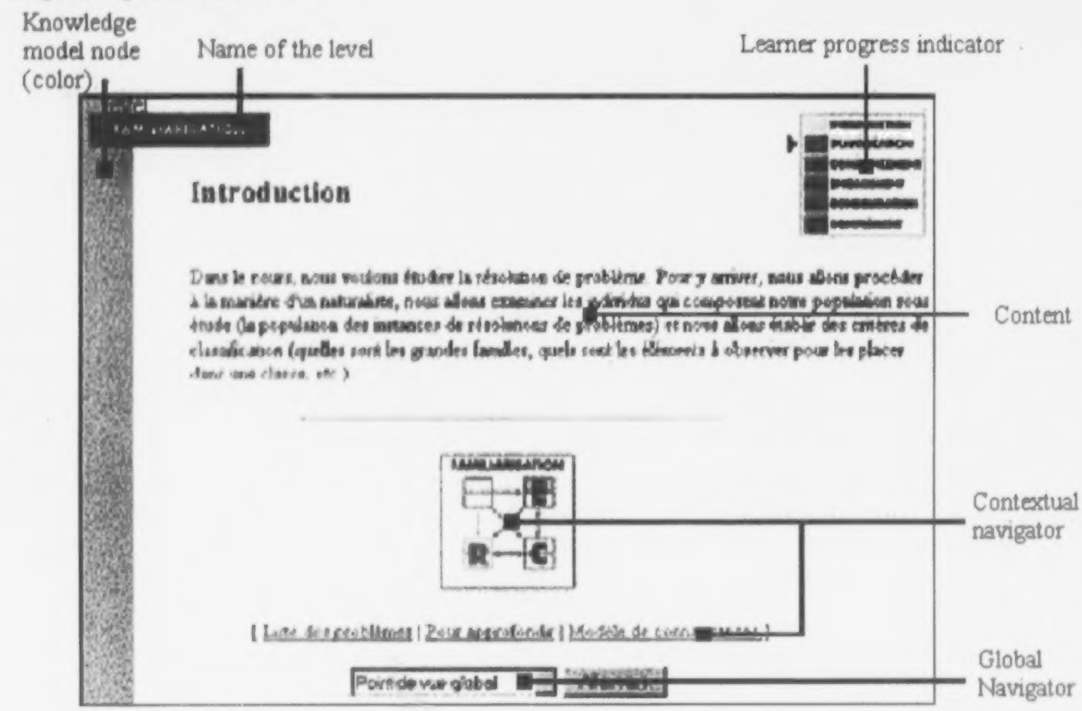


Figure 2: A view on a typical fragment.

4 Reusing and Maintaining Courseware: Agile LEs

In the preceding sections, we showed how the fractal structure of an LE can be used by a learner to adapt it to her cognitive style. In this section we show how such fractal LEs are *agile* with respect to their design, implementation and maintenance. Fragmentation into small pieces enables to tailor quickly them for specific purposes. The fractal design and the underlying grammar set up the formal grounds required to code procedures that generate LEs, extend them, manage updates, and maintain the site. Obviously, such properties have tremendous impact on costs.

The TEC6200 course addresses topics that are quite similar to those of INF6550. So we decided to reuse and extend the INF6550 LE to produce the TEC6200 one. First, we reused many fragments from the INF6550 LE. Since fragments were self-explaining and already contained information used by the contextual navigator, the process was easy and almost automatic. Then we extended the set of fragments of INF6550 LE by the addition of a new level respecting the self-similar structure on the one hand and by the addition of new peripheral fragments on the other hand. In the latter case, the grammar was extended to incorporate the new semantic knowledge components. We have also reused the interfaces of INF6550 LE with little modifications and 80 percent of INF6550 contents for two main fragments of TEC 6200 LE. Its two other fragments need complete processing on contents and visual aspects. We estimate that the reuse of INF6550 amounted to 28 percent saving on production, mainly because there was not been any prototype. On the other hand, improvements on the assistance infrastructure of TEC6200 will be injected into INF6550, so the savings will be more important.

Now that the INF6550 has been used by real learners, we have been able to verify that the site is effectively easy to maintain and to update in real-time. Fractal principles together with the grammar provide a principled way to update an LE and keep it a coherent. They enable one to code procedures that manage changes and maintain the site. To update an obsolete document, add information regarding a precise point, adding or retrieving an example,

BEST COPY AVAILABLE

we just need to put or retrieve the files on the site, since the appropriate hyperlinks are dynamically computed. The LE is thus reacting dynamically according to the information available on the site.

5 Conclusion

In this paper we outlined a framework we used to design and implement two learning environments on the WWW. Respecting the spirit of the WWW leads to highly fragmented documentation and activities. The fragments are assembled to produce an LE. Principles governing fractals self-similarity and 'infinite' decomposition guided the fragmentation of knowledge and ensured coherence of the LE's overall organization. Such coherence is compulsory to help the learner navigate through the knowledge and activities. Self-similarity also established a classification from which a grammar has then been derived. These LEs can ease the learning process and adapt to a given learner in the following ways:

- The fractal organization defines clues used for spatial orientation throughout the knowledge and the LE.
- The network structure and the fragmentation permit a learner to explore knowledge and activities according to its very own cognitive style. Its navigation can be interpreted in terms of the Myers-Briggs type indicator.
- Fragmentation and fractal design enable a learner to explore just the part of knowledge needed in a coherent way.
- Examples can be dynamically chosen according to the learner background.

These LEs are agile with respect to their design, implementation and updates in the following ways:

- *Reuse* : Fragmentation into small pieces enables to tailor rapidly an LE for specific needs, context and backgrounds.
- *Extensibility* : The fractal design sets up the rules for coherent extension of an LE.
- *Maintenance* : Fractal principles together with the grammar provide a principled way to update a LE and keep it a coherent. They enable one to code procedures that manage changes and maintain the site.

An advisor system using the learner's trace, the grammar and the Myers-Briggs type indicator is the next enhancement planned for these LEs.

6 References

[Chambreuil et al., 94] Chambreuil, Annie, Michel Chambreuil and Chihab Cherkouai, *Individualization Within a Multi-Agent Computer-Assisted*, Journal of Artificial Intelligence in Education, vol. 5, no 4, 1994, pp. 557-590.

[Giroux et al., 96] Giroux, Sylvain, Gilbert Paquette, François Pachet et Jean Girard, *EpiTalk, a Platform for Epiphyte Advisor Systems Dedicated to Both Individual and Collaborative Learning*, ITS'96 Third International Conference on Intelligent Tutoring Systems, Montreal, Canada, June 12-14 1996, Springer LNCS 1086, pp.363-371.

[Krebs, 85] Krebs, S., *Using the Myers-Briggs Type Indicator in Organizations*, Psychological Press, Inc., Palo Alto, CA, 1985.

[Newell and Simon, 76] Newell, A. and H.A. Simon, *Computer Science as Empirical Inquiry: Symbols and Search*, Communication of the ACM, vol. 19, no 3, 1976, pp. 113-126.

[Jonassen, 86] Jonassen, David H. *Hypertext Principles for Text and Courseware Design*, Educational Psychologist; vol. 21, no 4, pp. 269-292.

[Pierre & Hotte, 96] Pierre, Samuel and Richard Hotte, *Towards an Integrated Model for Supporting Collaborative Telelearning*, Annals of Telecommunications, vol. 51, no 5-6, 1996, pp. 272-287.

[Stepp, 96] Stepp, Ermel, *Fractal Frequently Asked Questions and Answers*, http://webpages.marshall.edu/~stepp/fractal-faq/faq.html.

MiMi: A Java Implementation of the MicroMint Scheme*

Vesna Hassler, Robert Bihlmeyer,
Michael Fischer, Manfred Hauswirth

Technical University of Vienna, Distributed Systems Group 184-1
Argentinierstr. 8/3rd floor, A-1040 Vienna, Austria
E-mail: hassler@infosys.tuwien.ac.at

Abstract: In this paper we describe an experimental implementation of the MicroMint micropayment scheme in Java. We apply this scheme to purchasing Web pages. A prerequisite was to accomplish this without having to change the code of either the Web server or the Web client. We discuss the implementation issues and security considerations. Our implementation requires the local protocol handler feature offered by Sun Microsystems' HotJava 1.0 browser.

Keywords: network security, Web security, Java security, electronic commerce, micropayment schemes

Introduction

The main motivation for introducing so-called *micropayment schemes* into electronic commerce protocols is that not all Internet commerce applications require transactions of large amounts of money. Accordingly, the security risks related to a single purchase are not so high. It is therefore rather expensive to deploy security mechanisms suitable for high security risks. For example, a typical charge for purchasing a Web page is one cent. Consequently, the only attack worth trying would be a large-scale forgery. Therefore micropayment schemes should be aimed at preventing large-scale attacks that would involve hundreds of thousands of purchases rather than at preventing a few losses in the range of one cent.

In a micropayment scheme, typical participants are a customer, a broker and a vendor. The customer buys *digital coins* from the broker and gives them to the vendor as payment for some service. The vendor returns coins to the broker in return for payment by other means (*redemption*).

In this paper we describe an implementation of a micropayment scheme (MicroMint [7]) called *MiMi*, applied for purchasing Web pages. In this setting the vendor is an information server that charges customers for accessing its Web pages. The server is implemented as a standalone Java application, but could also be implemented as an extension of a Web server (e.g. using the Java Servlet API [12]).

MicroMint

MicroMint [7] is a micropayment scheme intended for facilitating small purchases over the Internet. It offers low security, but is very fast because it makes no use of public-key cryptography. Its main advantages over other micropayment schemes [9] are as follows:

- it is off-line from the broker's point of view,
- it does not use either digital signatures or any other public-key scheme, and
- small-scale forgery attempts do not pay off.

At the beginning of each month the broker issues new coins. Unused coins are returned to the broker at the end of each month. Each coin is represented by k integer values (we use 32-bit integers) such that their hash

---

*This work was supported in part by a grant from Hewlett Packard Laboratories.

values (i.e. MD5 digests [6]) all have identical low-order n bits. This is called a *k-way collision*. Additionally, the c high-order bits of the hash value are specified by the broker, and are different for each month. For a detailed discussion of the MicroMint scheme see [7].

Java Security

The necessity for a sound security concept for the Java programming language results from the fact that most Java code is intended to be automatically downloaded across the network to run on a user's machine [13]. The main problem here, from the security point of view, is how to protect the user's host and data from being damaged by running malicious Java code. Due to the Java Virtual Machine [5] concept, Java code runs on all of the most popular platforms without recompilation. In other words, Java is an implementation of Web-based executable content.

The purpose of the Java security reference model [2] is the enforcement of Java language semantics [10] and a Java-enabled application's security policy. The Java Virtual Machine (JVM) enforces the Java language security features, like access modifiers for variables and methods. It calls the Class Loader in order to ensure that class names are mapped to class code in a proper way. JVM also provides the Bytecode Verifier to validate non-system classes. And finally, the Security Manager performs run-time checks on "dangerous" methods, like file read/write operations. Each Java-enabled browser uses its own version of the Security Manager. The Security Manager policy of most browsers is usually very restrictive. For example, applets cannot access local files at all. The new release of the HotJava browser (1.0 preBeta2) enables applets to gain different access permissions based on their digital signature [11].

The initial design of MiMi was intended to work with any Java-enabled browser. If a customer wished to purchase a Web page, an applet provided by the vendor would be downloaded by the customer's browser. This applet would take care of the communication between the customer and the vendor, i.e. its originating host, which is allowed by most browsers. However, since the vendor's applet cannot be trusted (most browsers' Security Managers do not allow non-local applets to access local resources at all), it could not read the coin(s) required for purchasing the requested page. Thus it would be necessary to have an additional, local applet that would communicate with the vendor's applet and operate on local files in which the coins and the security relevant information are stored. Unfortunately, inter-applet communication is not possible for applets with different security contexts (i.e. different Security Managers), so we had to abandon that solution.

HotJava 1.0 preBeta2 allows an applet to get access permissions for local files based on its certificate and digital signature. It is an extension of the Access Control Lists of Sun's Appletviewer [10]. This feature would allow a trusted digitally signed applet originating from the vendor to access the customer's wallet. A problem with this solution is that it might be necessary to repeatedly download the vendor's applet for each requested Web page. The vendor's applet should therefore stay resident in the browser and reactivate itself if the customer requested a new page from the same vendor.

In the current solution we don't use applets, but a locally installed protocol handler for HotJava. The client program defines a protocol called *MiMi*. The MiMi protocol handler is installed locally, so that it can get all permissions necessary to access local files without causing security problems.

An Overview of MiMi

MiMi comprises three Java applications (MMOrder, MMBroker and MMVendor), a protocol handler [15], as well as the HotJava 1.0 preBeta1 browser. The overall structure is depicted in Figure 2. The MiMi protocol handler enables the communication between the HotJava browser and the information server, i.e. MMVendor. Disadvantages of this approach are that the protocol handler has to be installed locally, and that this feature is currently not supported by browsers other than HotJava.

In our example setting MMVendor requires one digital coin for purchasing any of its Web pages. The customer can buy coins from MMBroker using the MMOrder application. The coins are stored in the customer's directory, in a file called *Wallet*. MMBroker mints coins and stores them in its own wallet file.

If the user wishes to access a MMVendor's page, s/he starts HotJava and types in a MiMi URL, like

mimi://host:port/dir/page.html

HotJava(tm): Connecting...
mimi://localhost:8081/homepage.html
Wallet

Figure 1: MiMi - Loading a Web page

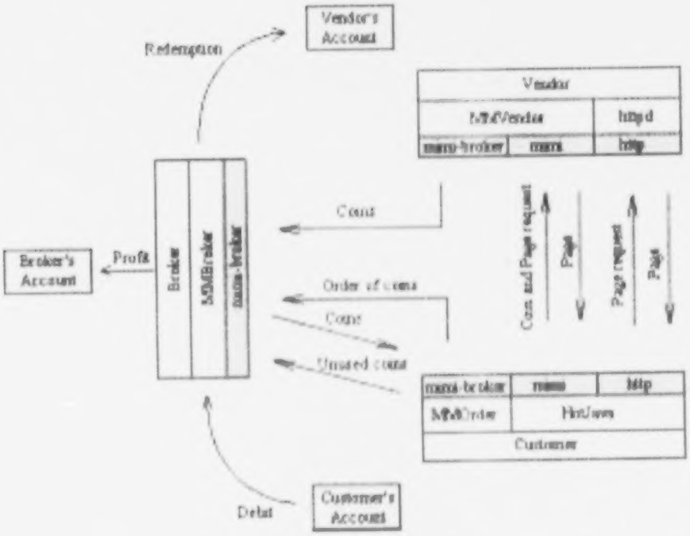

Vendor's Account
Redemption
Vendor
MiMiVendor
httpd
mimi-broker
mimi
http
Coins
Coins and Page request
Page
Page request
Page
Broker's Account
Profit
Broker
MiMiBroker
mimi-broker
Order of coins
Coins
Unused coins
mimi-broker
mimi
http
MiMiOrder
HotJava
Customer
Customer's Account
Debit

Figure 2: MiMi - An Overview

BEST COPY AVAILABLE

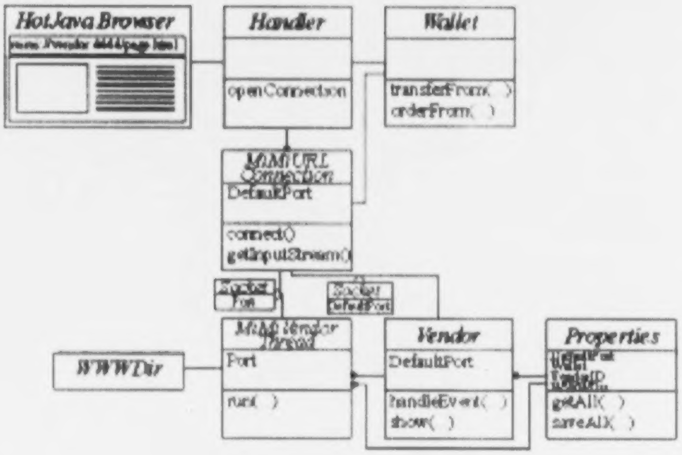


Figure 3: MiMi - The object model

(see Fig.1). When purchasing a Web page, the user is asked to pay one coin to MMVendor. MMVendor checks whether the coin is really a k-way collision, and whether it has already received it that month. If everything is correct, MMVendor accepts the coin and sends the requested page to the user. The user can view the page in his/her HotJava browser or, otherwise, the corresponding error message. At the end of each day MMVendor returns all collected coins to MMBroker. MMBroker checks each returned coin to verify whether it has been previously redeemed. For each valid coin MMBroker pays MMVendor a certain amount of money, e.g. one cent.

Some design issues

OMT model. In Fig.3 the OMT model [8] with the main vendor and customer classes is shown. For simplicity, we omitted some attributes and operations that are of little or no importance for this explanation.

How much to pay for a page? In the current MiMi implementation, one coin is required for one Web page. However, parts of a Web page (e.g. pictures) may be given as hyperlinks or as local links pointing to local files. If the reference is given as an hyperlink for the HTTP protocol, it is assumed to be public domain, so no additional coin is requested. If the reference is given as an hyperlink for the MiMi protocol, an additional coin is requested, i.e. a new window asking for a coin appears. If the customer does not want to pay for the "extra" pages, s/he can simply refuse further payments and download only the content the originally requested reference is pointing to.

MiMi security considerations

Customer-Broker. When purchasing coins from the broker, the customer must be sure that s/he is contacting the genuine one whose coins will be accepted as expected. In other words, the broker has to be authenticated. If the customer is authenticated, the broker can automatically withdraw the appropriate amount of real money from the customer's account, either locally at the broker or at the customer's bank. If the customer is not authenticated, s/he can anonymously order some coins from the broker and get them after having transferred the corresponding amount of real money to the broker's account. The coins must be transferred from the broker to the customer's wallet in a confidential way in order to prevent eavesdropping. The current version (February 1997) of MMOrder does not include authenticity and confidentiality, but we plan to implement these security services based on the SSL protocol [4]. Another possible solution is to use secure mail.

BEST COPY AVAILABLE

Broker-Vendor. Digital coins that the vendor has collected from the customers are redeemed by the broker that issued them. In order to prevent the man-in-the-middle attack it is recommendable to authenticate the broker, or at least use a long-term symmetric encryption key that would provide both weak authentication and confidentiality. Otherwise, using Web spoofing techniques [3] an attacker could masquerade as the broker, collect the coins from the vendor and redeem them at the genuine broker. In order to prevent eavesdropping, this exchange should be confidential. If the vendor is authenticated, the broker can automatically transfer the real money to its account. Otherwise, the broker could issue a digitally signed check and send it to the vendor in a confidential way. Here it is also be possible to use secure mail.

Customer-Vendor. The security problems that can arise by the customer-vendor communication are stealing of coins and stealing of Web pages. One of the design goals of MicroMint is to completely avoid public-key cryptography. However, it is recommendable to use a long-term symmetric encryption key between the customer and the vendor because it provides both weak authentication and confidentiality. This method would protect against stealing of both coins and Web pages. For each exchange of the long-term key the vendor should be authenticated using some strong authentication protocol. There are also other techniques to prevent stealing of coins proposed by the authors of MicroMint, like user-specific or vendor-specific coins [7]. If the [non-specific] coins are sent in cleartext, an attacker could use Web spoofing techniques [3] to collect coins and send in return fake Web pages. However, this attacker would have to provide Web pages that in the long run look similar to the genuine pages, to a large number of customers. Otherwise, this attack would not pay. If the vendor's Web pages are sent in cleartext, an attacker could collect them, become a vendor him/herself and sell the stolen pages. However, this would be revealed pretty soon, by the genuine vendor or by an honest customer. Moreover, if the contents of the Web pages change on a daily basis (like newspapers), this type of attack does not pay at all.

Conclusions

In this paper we presented a simple solution for applying the MicroMint scheme to purchasing Web pages. At the moment this solution works with Sun's HotJava browser only, because we use one of its advanced features (locally installed protocol handler). We hope that in the near future this feature will be offered by other browsers as well, and that it will be possible to dynamically load the protocol handler.

The new release of HotJava 1.0 (preBeta2) enables applet authentication, so that an applet can access the local environment if it is digitally signed and if its originator has a proper certificate. Having this feature in Java-enabled Web browsers would make it possible for the protocol handler to work with applets loaded over the network, even without an integrated protocol handler support [14].

If the protocol handler could also be loaded dynamically, it would have to undergo strict security checks. This is most probably the reason why the new HotJava release (1.0 preBeta2) still does not allow dynamically loaded protocol handlers, although it was expected; this feature would namely require a security concept, similar to applets.

References

[1] Berners-Lee, T., R. Fielding, H. Frystyk, *Hyptertext Transfer Protocol – HTTP/1.0*, Request For Comments, RFC 1945, May 1996, URL:<ftp//ds.internic.net/rfc/rfc1945.txt>

[2] Erdos, M., B. Hartman, M. Mueller, *Security Reference Model for the Java Developer's Kit 1.0.2*, Nov 1996, URL:<http//www.javasoft.com/security/SRM.html>

[3] Felten, E.W., D. Balfanz, D. Dean, D.S. Wallach, *Web Spoofing: An Internet Con Game*, Technical Report 540-96, Department of Computer Science, Princeton University, 1996
URL:<http//www.cs.princeton.edu/sip/pub/
spoofingDocumentWithLongUntypeableName.html>

[4] Freier, A.O., P. Karlton, P.C. Kocher, *The SSL Protocol, Version 3.0*, Internet Draft, March 1996, URL:<ftp//ietf.cnri.reston.va.us/internet-drafts/draft-freier-ssl-version3-01.txt>

[5] Lindholm, T., F. Yellin, *The Java Virtual Machine Specification*, Addison-Wesley: Tha Java Series, 1996

[6] Rivest, R.L., *The MD5 message-digest algorithm*, Internet Requests for Comments, RFC 1321, April 1992

[7] Rivest, R.L., A. Shamir, *PayWord and MicroMint: Two simple micropayment schemes*, also presented at the RSA '96 conference, URL:<http//theory.lcs.mit.edu/ rivest/RivestShamir-mpay.ps>,

[8] Rumbaugh, J., M. Blaha, W. Premberlani, F. Eddy, W. Lorensen, *Object-Oriented Modeling and Design*, Prentice Hall, 1991

[9] Hallam-Baker, P.M., *Electronic Payment Schemes*, URL:<http://www.w3.org/pub/WWW/Payments/roadmap.html>

[10] JavaSoft, *Frequently Asked Questions: Applet Security*, URL:<http://java.sun.com/sfaq/index.html>

[11] JavaSoft, *Java Security: New!*, URL:<http://www.javasoft.com/security/index.html>

[12] JavaSoft, *The Java Servlet Development Kit*, URL:<http://jeeves.javasoft.com/products/java-server/sdk>

[13] McGraw, G., E.W. Felten, *Java Security. Hostile Applets, Holes and Antidotes*, John Wiley & Sons, Inc., 1997

[14] Niemeyer, P., J. Peck, *Exploring Java*, O'Reilly & Associates, Inc., 1996

[15] Sun Microsystems, Inc., *HotJava 1.0 preBeta2: User's Guide: Installing a Local Protocol Handler*, 1997

A Toolset for Personal Information Management

Daniel J. Helm, Raymond J. D'Amore

Digital Libraries Department
The MITRE Corporation
1820 Dolley Madison Blvd., McLean, VA 22102-3481, USA
{dhelm, rdamore}@mitre.org

Abstract: We have developed an integrated set of advanced processing tools enabling a user to collect, organize, and summarize documents collected from both internal and external data repositories. The tools enable a user to retrieve and save documents to personal folders, and to then organize the collected information into related piles using clustering techniques. The documents can be collected while browsing information repositories or with the aid of an integrated search agent. Document summarization tools are also provided to extract key concepts from individual documents through statistical and natural language processing techniques. Together, these tools provide a means for a user to collect, organize, and assimilate the potentially large quantities of retrieved information returned from existing information retrieval systems as a basis for more effective analysis and discovery .

1 Introduction

Over the past few years, the area of Internet-based resource discovery has gone through a number changes. During the early 1990's, there was little software support for the user to locate and discover useful resources. Most directories were manually constructed, and provided and index to only a small fraction of the information space which in general consisted of anonymous FTP sites. Around 1991, tools such as Archie [Emtage and Deutsch] was introduced to provide a global searchable index to the information space. Archie provided the first large-scale index to the Internet, but since its index included only file names, only a very limited search capability was provided. Unless files were named in such a way as to reflect their contents, they could not be found by Archive. Around the same time WAIS [Kahle and Medlar] was introduced, which provided a full-text search capability to document collections. WAIS divides its indices among the servers that provide information, rather than using one centralized index as with Archie. This does not enable global searches to be performed, but requires that a user access a central index of servers from which individual WAIS servers can then be selected and queried. A number of query routing extensions have been made to WAIS [Gravano et at. 93, Sheldon et al. 94] enabling a single query to be issued and broadcast to multiple WAIS servers, where results from multiple servers are merged and returned to the user.

With the introduction of the Web, a number of powerful search systems have been introduced in recent years. These systems typically have a "spider" component that collects documents from the Internet which then feeds an indexing subsystem to provide a full-text search capability to the collected information content. Current search systems (e.g., Altivista, Lycos, Excite) provide varying degrees of indexing coverage and provide search options ranging from "simple" keyword-based search to more advanced options including support for Boolean operators and natural language-like queries. Most search systems create a centralized (usually replicated) document index, and in some instances index a very significant portion of the Web. Recently, a number of search agents [Genesereth and Ketchpel] have been introduced to provide a higher-level interface to existing search engines. Search agents collect queries from users, query multiple predefined search engines, and then merge and return results to users. Agents can benefit the user by not requiring that s/he contact, in a serial manner, multiple search engines to obtain relevant information. A number of other systems (e.g., Yahoo) were also introduced to primarily provide a browse interface to information organized into a taxonomy of topical areas. Information in these systems tends to be manually organized and does not contain the volume of information indexed by typical full-text search systems.

One problem with most existing search systems is that they return large numbers of documents which are often only denoted by subject lines and possibly simple "abstracts" that include the first few lines of text from each document. What is currently lacking with most of these systems are tools that provide a means to organize and assimilate the potentially large quantities of retrieved information as a basis for more effective analysis and discovery. Our preliminary assessment has indicated the need to balance available information retrieval and classification capabilities with a new generation of highly efficient post retrieval analysis tools for extracting, organizing, and visualizing information within extensive results sets. These back-end processing tools will be user accessible "on demand" through an object oriented interface to provide users with methods for maintaining personal views of large, heterogeneous information spaces.

To address this problem, we have developed an integrated set of advanced processing tools enabling a user to collect, organize, and summarize documents collected from both internal and external data repositories. The tools enable a user to retrieve and save documents to personal folders, and to then organize the collected information into related piles using clustering techniques. The documents can be collected while browsing information repositories or with the aid of an integrated search agent. Document summarization tools are also provided to extract key concepts from individual documents through statistical and natural language processing (NLP) techniques. Together, these tools provide a means for a user to collect, organize, and assimilate the potentially large quantities of retrieved information returned from existing information retrieval systems as a basis for more effective analysis and discovery.

2 Personal Information Management Toolset

The personal information management toolset currently consists of the following subsystems: information collection, information organization, and document summarization. One key goal of our system is to provide a set of modular information collection and management services that can be easily extended and/or replaced. These services are made available through a framework that integrates our own software with publicly and commercially available software.

2.1 Information Collection

Information collection is achieved using two mechanisms: a search agent interface and a manual browsing technique. The browse method enables a user to save off to personal folders selected documents while perusing information repositories. This capability enables one to save either the currently displayed document, or also all top level documents referenced (via URLs) in the current document.

The search agent interface receives a simple keyword based query from the user, queries a predefined set of search engines, merges the retrieved document lists, and stores the results in an existing or dynamically generated folder. The search agent can be invoked interactively or it can be scheduled to run in the background at periodic time intervals. When invoked in the background, the user will be alerted via e-mail when any new documents arrive.

2.2 Information Organization

The current method used to organize information includes hierarchical folder management in conjunction with robust, adaptable, and efficient clustering methods which are extensions to those developed in our previously developed system [Helm et al.]. The folder management scheme enables a user to, manually or automatically via a search agent, assign documents into topical folders. The clustering tools provide a finer grain topical space than that provided through folders alone, by automatically generating dynamically defined topical subclasses within a folder. The clustering algorithm utilized supports a pre-cluster analysis stage and post-cluster refinement stage. The pre-cluster stage can use various information compression and sampling schemes for reducing the size of the problem (i.e., reduce the computational costs.) The post-cluster stage can be used to prune clusters and otherwise modify cluster content to either improve cluster effectiveness directly, or to possibly stage for another run when the system is setup for multi-pass clustering. The actual clustering algorithm is parameterized to allow for modification of the similarity thresholds used to determine document assignment into topical classes, the rules for managing multi-class assignments, and the method used to compute centroids dynamically as the cluster changes. The clustering algorithm also automatically generates cluster summaries and cluster labels to support user review. Fig. 1 shows the major processing steps utilized in our clustering algorithm. The two cycles shown above the figure imply that adjacent processing steps can iterate and utilize data feedback to enable the clustering algorithm to be more dynamic and adaptable.

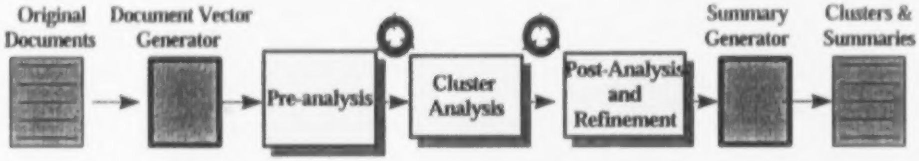


Figure 1. Document Clustering Processing

2.3 Document Summarization

We provide a number of statistical and NLP-based document summarization tools to extract key items and sentences from documents enabling a user to more quickly determine general themes and topics described in a document. We currently provide three summarization tools: query independent summarization, query dependent summarization, and entity-based summarization.

For our non entity-based summarization tools, we use an extended version of a part-of-speech (POS) tagger developed by Eric Brill [Brill] to assign POS tags to words in a document. Depending on the type of summarization performed (query dependent or independent) we perform the following. For query independent summarization, we extract all noun phrases from the POS tagged document, and then perform statistics on the resulting phrases to identify those that occur with the highest frequency. We also will identify key sentences which contain the greatest number of key noun phrases. The resulting list of phrases and sentences are used as a summary for the document. For query dependent summarization, we simply extract all sentences and phrases that contain any query terms. The POS information is used to extract noun phrases that contain one or more query terms. The resulting phrases and sentences are then statistically scored and used for a document summary. To improve the summarization process, we utilize a stemming algorithm [Frakes and Baeza-Yates] to expand term associations. Fig. 2 shows the document summarization processing steps for both query-dependent and query-independent summarization techniques.

A third type of summarization we provide utilizes an entity tagger developed by IsoQuest [NameTag]. This software enables us to extract entities such as person, place, and organization names from a document. It provides a richer phrase level summary than simple noun phrases alone by assigning type classifiers to multi-word phrases.

BEST COPY AVAILABLE

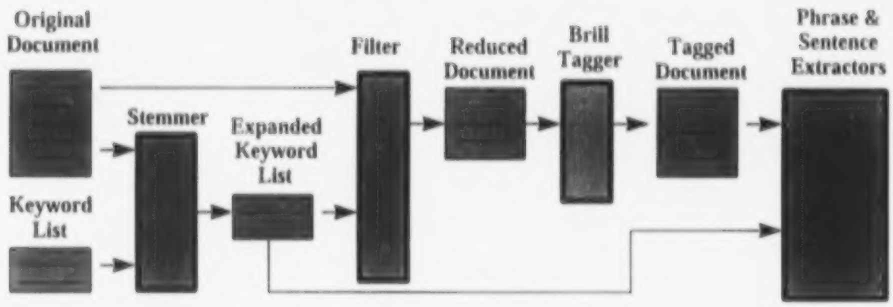


Figure 2. Document Summarization Processing

3 Prototype System

We have developed a prototype system to evaluate our ideas as well as to serve as a testbed for future extensions. The prototype is client/server based, and currently utilizes a Sun SPARCstation 10 running Solaris 2.4 for all basic processing including folder management, clustering, and summarization. The client interface is any Web browser supporting JavaScript (e.g., Netscape Navigator). All server-based software is written in C and Perl.

Fig. 3 depicts two different view types that are available for displaying the contents of a folder: document- and cluster-based views. A user can click on a folder from the left frame of a window to view all documents that had been collected into the selected folder, where the results will appear in the right frame in the window (top screen). From the document list screen, a user can click on the "List clusters" button to view the dynamically generated clusters for the list of documents (bottom-left screen). Each cluster is followed by the top discriminating phrases for documents in the cluster. We cluster using keyword-based document vectors, but then as part of the cluster summary generation process, we use POS tagging to extract these noun phrases from the original documents to provide more useful summaries. From the cluster list screen, a cluster can be selected to display all documents assigned to the cluster (bottom-right screen). The number to the left of each document is a cluster similarity score.

BEST COPY AVAILABLE

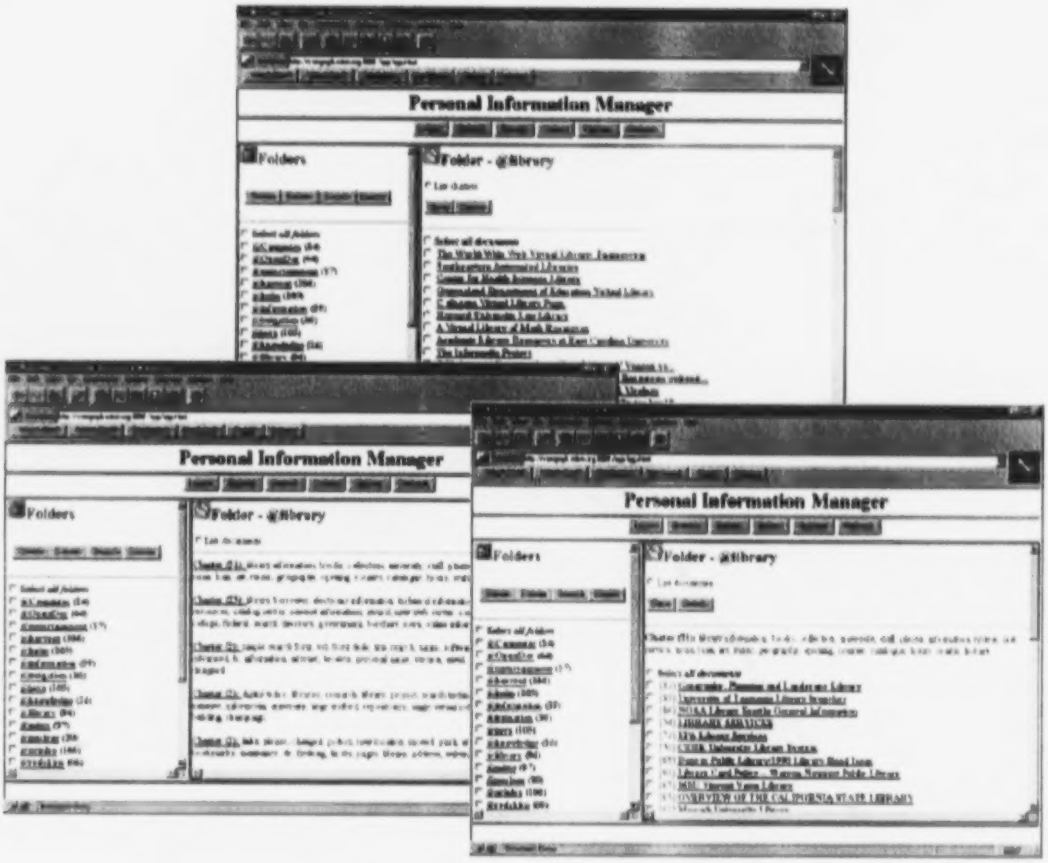

Personal Information Manager
Folders
Folder - @library
Personal Information Manager
Folders
Folder - @library
Personal Information Manager
Folders
Folder - @library

Figure 3. Document and Cluster List Screens

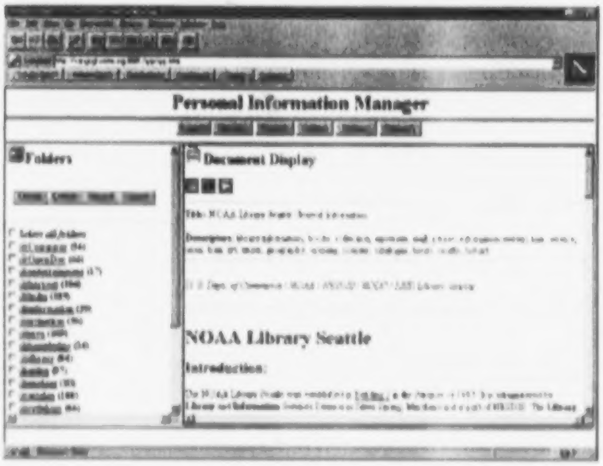

Personal Information Manager
Folders
Document Display
NOAA Library Seattle
Introduction:

Figure 4. Document Display Screen

BEST COPY AVAILABLE

From either the document list or cluster screen, a document can be selected and displayed as shown in Fig. 4. From this screen, three types of document summaries are available via three icons shown at the top of the document display frame: query independent, query dependent, and entity-tagged. Three sample document summaries for the document are shown in Fig. 5, where the top screen shows a query independent summary, the bottom-left screen a query dependent summary, and the bottom-right screen a entity-tagged summary.

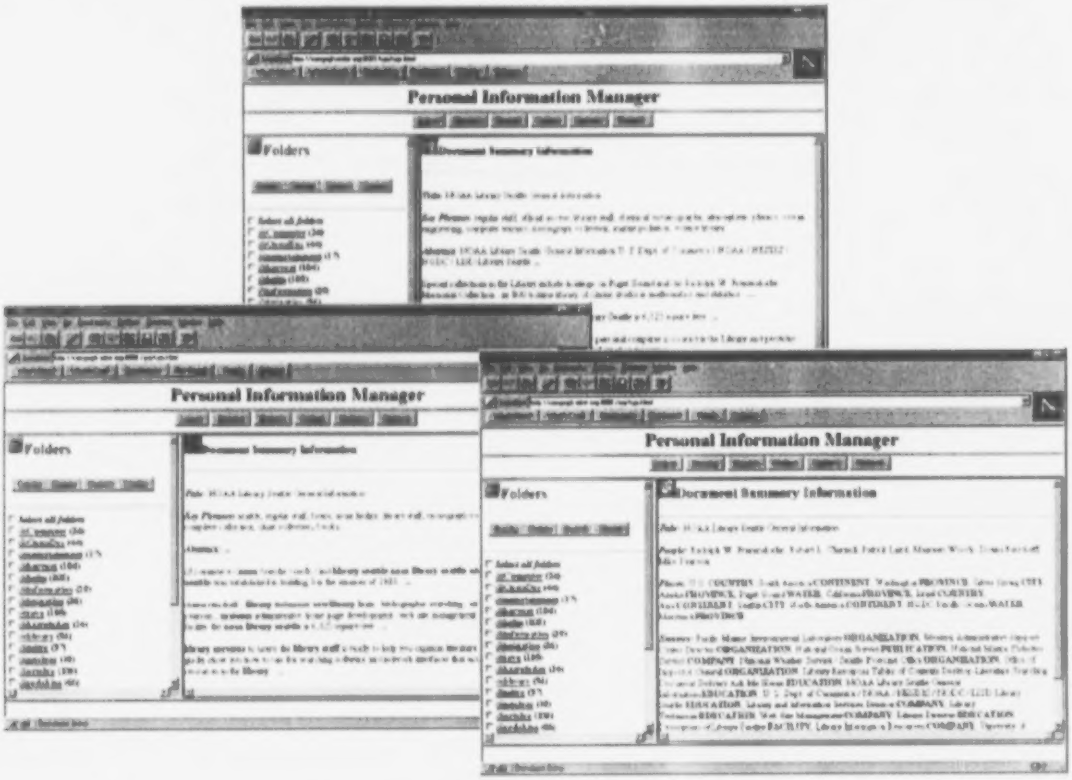

Personal Information Manager
Folders
Document Summary Information
Personal Information Manager
Folders
Document Summary Information
Personal Information Manager
Folders
Document Summary Information

Figure 5. Document Summarization Screens

4 Conclusion and Future Work

This paper has described a baseline personal information management system that provides tools to collect, organize, and assimilate the potentially large quantities of retrieved information returned from existing information retrieval systems as a basis for more effective analysis and discovery. The system is currently being evaluated by a user group within the MITRE Corporation to determine system effectiveness and to provide feedback as a basis for making extensions.

Some planned enhancements include extending the document summarization methods to groups of documents. Currently, our summarization tools operate on individual documents. We would like to summarize groups of documents (as with clustering) within personal folders. Another possible enhancement includes support for visualization. Currently the presentation of the system's results are text-based. We would like to extend the presentation to allow for more graphical displays, in particular for viewing clustering results. This may provide a more intuitive way to view large document spaces. We may also extend the search engine interface to provide support for heterogeneous systems. Currently the search engine interface can access search engines only via URLs. We may extend the agent to enable it to query diverse repositories such as database systems.

BEST COPY AVAILABLE

References

[Brill] E. Brill. A Corpus-Based Approach to Language Learning. Ph.D. thesis, Department of Computer and Information Science, University of Pennsylvania, 1993.

[Emtage and Deutsch] A. Emtage and P. Deutsch. Archie - An Electronic Directory Service for the Internet. Proceedings of the USENIX Winter Conference, pages 93-110, January 1992.

[Frakes and Baeza-Yates] W. B. Frakes and R. Baeza-Yates. Information Retrieval: Data Structures and Algorithms. Prentice-Hall, 1992 ISBN: 0-13-463837-9

[Genesereth and Ketchpel] M. R. Generereth and S. P. Ketchpel. Software Agents. Communications of the ACM, 37(7), pages, 48-53, July 1994.

[Gravano et al.] L. Gravano, H. Garcia-Molina, and A. Tomasic. The Efficacy of GlOSS for the Text Database Discovery Problem. Technical Report STAN-CS-TN-93-002, Stanford University, November 1993.

[Helm et al.] D. J. Helm, R. J. D'Amore, and P. Yan. MIDS: A Framework for Information Organization and Discovery. Journal of Network and Computer Applications, 19(4), pages 381-394, Academic Press, November 1996.

[Kahle and Medlar] B. Kahle and A. Medlar. An Information System for Corporate Users: Wide Area Information Servers. ConneXions - The Interoperability Report, 5(11), pages 2-9, Interop, Inc., November 1991.

[NameTag] see URL: http://www.isoquest.com.

[Sheldon et al.] M. A. Sheldon, A. Duda, R. Weiss, J. W. O'Toole, Jr, and D. K. Gifford. A Content Routing System for Distributed Information Servers. Proceedings of Fourth International Conference on Extending Database Technology, March 1994.

Acknowledgments

We would like to thank Thomas E. Bagby, Steven H. Kerchner, Gary L. Valley, and Puck-Fai Yan who contributed to the implementation of the system.

World Wide Web Presentations for Volatile Hypermedia Database Output

Geert-Jan Houben[*] and Paul De Bra[**]
Department of Computing Science
Eindhoven University of Technology
PO Box 513
NL 5600 MB Eindhoven
The Netherlands
{houben,debra}@win.tue.nl

[*] Geert-Jan Houben is also affiliated with Origin.
[**] Paul De Bra is also affiliated with the University of Antwerp, and with the "Centrum voor Wiskunde en Informatica" in Amsterdam.

Abstract: The use of databases in current information systems is changing rapidly. The information contained in the database shows less structure, and contains multi-media objects and free-text, as well as hypertext links. To represent the information from a database a hypermedia platform such as World Wide Web often appears to be a right choice. Typical examples of such hypermedia applications are employee databases that include both administrative and personal information, museum databases that offer guided tours as well as query facilities for their collection, geographic information systems as used in tourist applications, and mail-order catalogs and services. We claim that hypermedia applications can help to represent the output of such databases.

The Relationship Management Methodology (RMM) [Isakowitz et al. 1995] is a hypermedia design method developed specifically for generating hypermedia navigation and presentations of database information. RMM helps in designing presentations for (whole) information objects contained in a database, and in generating navigational structures such as indexes and guided tours.

A significant part of the database information that an application must present, is of a volatile nature. Most of the output of the database is the result of queries. Such output does not have a predefined structure for which a representation could be designed using RMM. The desired presentation may also depend on the size of the query-answer (i.e. on the number of objects in the answer). Since RMM is a design method consisting of steps performed by human designers, it does not provide a means for generating presentation structures for answers to arbitrary queries.

In this paper we present an approach for extending the principles of RMM in order to generate hypermedia (HTML) presentations and navigational structures for query answers. First, we propose heuristics that are based on Relationship Management Design Model (RMDM) structures generated through RMM to provide presentational and navigational cues in a database query. Moreover, we give ways to override these defaults through extensions to the SQL query language.

Keywords: hypermedia design, presentation of query results, generation of navigation, generation of presentation

1. Introduction

The way in which databases are used in current information systems often differs significantly from the way we are used to. The information contained in the database is less strictly structured than in traditional (administrative) information systems. The term semi-structured data is often used to describe information that contains free-text components and multi-media objects. The use of a hypermedia platform such as World Wide Web can help to represent the less structured information. Typical examples of such applications are employee databases, museum databases, geographic information systems, and mail-order catalogs and services. The design and construction of such hypermedia applications for database information is the main subject of this paper.

Designing and constructing a hypermedia application involves the representation of relationships among information objects. The Relationship Management Methodology (RMM) [Isakowitz et al. 1995] is based on the Entity-Relationship model [Elmasri et al. 1990] and on HDM [Garzotto et al. 1991]. It proposes a methodology to support the design and construction of an application by suggesting a translation from an E-R design to a navigational design with RMDM (RMM's data model) constructs. The methodology is augmented with RMCase [Díaz et al. 1995], a CASE tool that provides not only a graphical interface for developing RMDM constructs, but also an interface for designing the presentation of an RMDM construct as an HTML document. Note that RMM is intended for multimedia databases that combine free (or almost free) text, images and possibly also other information. For "administrative" data the generation of a hypermedia (really hypertext) representation is much easier, as demonstrated by [Pönighaus et al. 1996].

RMM includes a translation from E-R models to RMDM models. In RMDM E-R relationships are replaced by navigational structures such as indexes and guided tours. Entities, especially large entities, are divided into slices for better presentation. Slices are groups of attributes which belong together semantically and which are thus presented simultaneously. For each entity there is a head slice, presenting the most essential information, and containing (hypertext) links to the other slices.

The subject of our interest is the presentation (through WWW browsers) of volatile database output. This volatile information is the result of queries executed by the database. When we look at the information contained in a query result, that information depends on two aspects:

- the definition of the data in the database, i.e. the (static) data structures that underly the volatile information;

- the specification of the exact query, i.e. the dynamically determined aspects that are necessarily attached to the volatile information produced by the specific query.

We find that RMM (and RMCase) do not help in the case of volatile data. The main problem is that the dynamically generated structure of a query result cannot be (trivially) translated into a hypermedia presentation.

- The RMM approach can be applied to applications of a fairly stable nature. For the hypermedia applications based on more volatile information the RMM guidelines fall short since much of the information about the database, and especially the dynamically determined aspects of the query results, is not explicitly contained in the associated data model. So, RMM can help to guide a designer to find a proper representation by using the data definition. However, the exact specification of the query needs to be considered as well to produce a fully suitable hypermedia representation.

- Also, since the number of possible data structures of query results is very large it is not feasible to create the navigational structure and the HTML presentation through RMCase for each possible query. In [Díaz et al. 1995] the designers of RMCase acknowledge that the creation of hypermedia representations for database queries is a complex task, not handled by RMCase.

The core of our proposal in this paper is to use as much as possible the presentation aspects that are related to the data definition: if possible, use the presentation of the data structures underlying the query information. However, in addition to the use of the data structures this paper includes general guidelines to present the dynamic, query dependent information in hypermedia format.

In order to be able to do so, we must consider the information from the database in both the data definition and the data manipulation perspective. We will use SQL as the language to specify data definition and manipulation. We give a translation from SQL to RMDM while using the existing translation from E-R to RMDM. Thus, we present an extension to the RMM methodology that bridges the gap between SQL and RMDM. The extension to RMM is twofold:

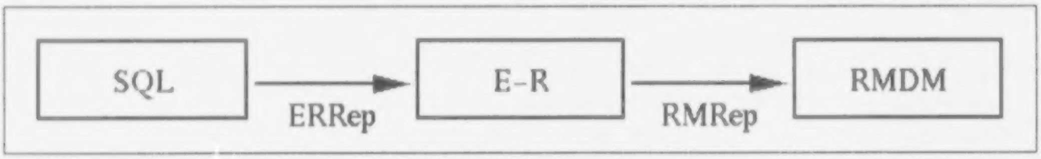
1. We show how the data definition component of the database can be used to propose a default representation in RMDM structures for any query result.
2. We suggest an extension to SQL (to be interpreted by a preprocessor) to offer the user the possibility to include representation information (i.e. navigation and presentation) in a query.

2. RMM and SQL

As we have seen in the previous section, RMM offers a translation from E-R to RMDM. This translation is elegantly presented in [Isakowitz et al. 1995]. For the purpose of representing the structural database information by means of a hypermedia format, this translation suffices. There we can limit ourselves to the data definition involved. In order to be able to translate queries (dynamically determined information) we must take the data manipulation aspects into account as well. Therefore, we choose to use SQL as the vehicle for data definition and manipulation. This implies that SQL is the source platform for the translation to RMDM.

We need to consider another translation, the one from SQL to E-R. In practice an E-R diagram needs to be translated to a set of relations (in the relational database model) in order to obtain an implementation. This translation from the E-R model to the relational model is straightforward (and often done according to a standard method). It holds that the inverse translation is equally simple: it is this translation (from the relational model to the E-R model) that we must consider to complete the global translation from SQL (the relational model) to RMDM. We assume here that we know how entities and relationships are modeled by the given database relations.

These two translations (called ERRep and RMRep in the next figure) build the mechanism to translate the major data definition elements into RMDM representations. For each relation in the database, specified in the SQL data definition part, a representation in RMDM is generated by applying the translations ERRep and RMRep.



For the sake of translating data definition aspects, this approach suffices. We can not apply this approach to the data manipulation aspects. When we consider the volatile information specified by SQL queries, the translation through E-R can not be used as an intermediate step in all circumstances. In general, this E-R intermediate is not available for data manipulation: we then must be able to use a direct translation from the SQL context to RMDM.

3. Hypermedia Representation and Data Definition

To obtain a generally applicable translation mechanism, we have found it effective to use as much information as possible from the data definition part of the database. By extracting as much information as possible from the data dictionary, the core of the representation in hypermedia format (RMDM) should become available. Thus, the representation that we propose for volatile database output is based on a default representation determined by the data definition aspects. Subsequently, the default (standard) representation is adjusted to the specific details in the data manipulation aspects, i.e. the dynamically determined aspects of the information.

To be able to deduce an elegant hypermedia presentation, details should be included in the data definition that concern the user preferences for the representation aspects. We present here some items that can be specified in the user preferences.

A number of issues concern the relations involved:

- The user can specify for every relation how the set of records (instances) is presented: one per page, all on the same page, or a fixed number of records presented on one page. The user should determine how many records are shown on one page. This issue gives the user the possibility to adjust the presentation to the semantics and the size of the records.

- Related to the issue of the presentation of records on pages, is the issue of the connection between those pages: the user must specify the access structure of the relation, and thus the way in which the different pages are accessed. In the data definition the user specifies for every relation the standard access structure, like index or guided tour. In this way the user can express the access routine that best suits the way in which end-users want to access the different pages.

- The above two issues must be determined for the presentation of the relation itself, but also for query results that are a (small) subset of the entire relation. As the presentation of a query result might differ from the general presentation, we give the user the possibility to add additional information related to the size of the query result. The user can specify a different number of records per page and a different access structure, if the number of records to show is less than a given number. The idea is that it could be wise to present a small number of records on one page, while presenting large numbers of records through an index: since this can differ for subsets of a relation, we want to give the user the possibility to specify different presentations for different query results.

RMM associates a set of slices with every relation. Some presentation issues concern the slices:

- In order to be able to access the different slices, an access structure for the slices of a relation must be specified. It must be determined how slices are accessed from the head slice. In this way the different semantics attached to the different slices can be expressed.
- Besides the access structure for the slices, the user must specify for every slice which presentation (layout) should be used. This implies that the user determines how a specific slice is presented in hypermedia format.
- In the same way as for relations the presentation can depend on the size of the (computed) relation; hence we allow the user to specify how the slice presentation and access structure depends on the number of slices (records) to show. This includes the number of slices (records) presented on one page and the access structure for those pages.

As we will see later, the presentation of database output (query results) implies the creation of new slices that contain (only) the information explicitly asked for in a query. For these newly created (volatile) slices presentation aspects can be defined in the data definition:

- The user should have the possibility to specify beforehand the default presentation (layout) to be used for new automatically generated slices. It appears that for a number of aspects the user can specify a framework for presentation beforehand. This framework can later be used to determine the actual presentation of a specific query result. Of course, this framework can be defined in general (for all relations), but we allow the user to specify (override) a specific framework for a given relation.
- As before for the relations and (structural) slices, for the other presentation issues (access structure and records per page) a default can be proposed for representation of the volatile slices. Again, this default can be overridden in the case of specific query, but one can expect that per relation a default can be determined beforehand for its volatile slices.

The nature of this approach is that during the data definition a choice is made between for example index and guided tour, based on the semantics and syntax (layout) of the items (entities or relationships) involved. Moreover, based on the same semantics and syntax the most elegant combination of records on separate pages is chosen. The general idea in this representation is that most presentation issues are settled beforehand, thus giving a default (framework) that can act as the basis for determining the exact presentation during data manipulation.

We summarize the above by stating the presentation aspects specified by the user on the basis of the data definition:

the number of records per page	(for the complete relation)
the number of records per page	(for subsets)
the access structure for connecting pages	(for the complete relation)
the access structure for connecting pages	(for subsets)
the presentation of slices	(for the complete relation)
the presentation of slices	(for subsets)
the access structure for slices of one relation	(for the complete relation)
the access structure for slices of one relation	(for subsets)
the presentation of newly created, volatile slices	(dependent on the size)
the access structure for volatile slices	(dependent on the size)

4. Hypermedia Representation and Data Manipulation

In this section we concentrate on the routine that is used to produce a hypermedia representation during data manipulation. The prime idea behind this routine is that the user is offered a predefined, default representation, based on the presentation aspects included in the data definition.

The input for the routine during data manipulation is an SQL query. Our routine to produce a proposal for the representation of the results of SQL queries includes three main steps:

1. The exact query specification is considered to see what the result of the query can borrow from the RMDM representation issues of the relations closely related to that result: these issues are available in the data (definition) dictionary and act as the default to start from.
2. The default representation deduced from the data definition is adjusted in correspondence with the number of records found in the result.
3. The proposed representation is overridden and adjusted, if the query specification contains elements that explicitly ask for a given representation.

In this section we now address the first two steps: the third step is covered in the next section.

To illustrate the use of predefined representation aspects we must start the data manipulation from a given SQL query. Let us assume that this query looks like

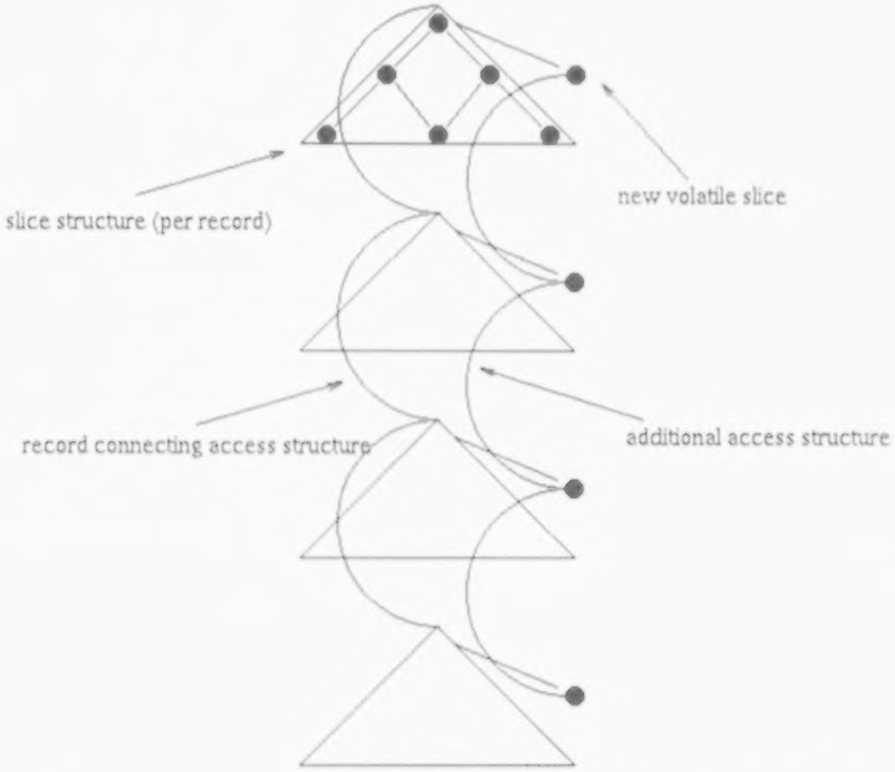
```
SELECT <Attributes>
FROM <Relations>
WHERE <Condition>
```

The outline of the routine distinguishes three cases.

4.1 One Slice, One Relation

In this first case the attributes in <Attributes> all belong to one slice, say S, of a relation, say R. Then we use the principle that if the user specifically asks for attributes from one slice, this slice should be offered (directly accessible) to the user.

As far as the access structure for these records (slices) is concerned, we choose to borrow the representation of R (from data definition): so, we use slice S as the entry slice to the relevant records, which are accessible to their default access structure (from data definition). At the same time, the head slices remain also connected through that similar access structure.


slice structure (per record)
new volatile slice
record connecting access structure
additional access structure

entry slice	slices	presentation	record access structure
the slice from the	as in data	as in data	as in data definition

BEST COPY AVAILABLE

SELECT clause	definition	definition	(between head slices and SELECT-slices)

We assume here that in general every slice is designed in such a way that the user is satisfied with the presentation of the entire slice (and does not feel the need to limit the presented data). So, the fact that the user asks (in the SELECT clause) for some of the attributes from a given slice, is interpreted as asking for that (whole) slice.

4.2 Multiple Slices, One Relation

Another possibility is that the attributes in <Attributes> belong to multiple slices of one relation R and <Relations> only contains R. We then use the idea that if the user asks for attributes from multiple slices, the user should be offered the relevant records through their head slices, with additional access to a new (volatile) slice with exactly the attributes from <Attributes>. The new slice gets a default name, e.g. NEWSLICE. So, the representation looks like the representation of R for the relevant records, with the head slice as its entry slice and one additional (volatile) slice containing all the attributes of <Attributes>.

As far as the access structure between records is concerned, we choose to use the same principle as above. This means that the default access structure between records is borrowed from the data definition, and it is also implemented to connect the new volatile slices. Just as in the previous case, we obtain two record connecting access structures: one between the head slices, and one specially for the slices implied in the query.

To the access structure between slices (within a record) we necessarily add two links: one from the head slice to the new volatile slice, and one link in the opposite direction. These links get default names, e.g. HEAD-NEW and NEW-HEAD.

entry slice	**slices**	**slice access structure**	**presentation**	**record access structure**
head slice	as in data definition plus a volatile one	augmented with two new links	as in data definition	as in data definition (between head slices and volatile slices)

The principle behind this proposal is that the result of the query models a set of records and that with the SQL query the user is searching for relevant records: therefore, we propose to access these relevant records through their standard head slice.

Just as in the previous case, if the user does not want to see a new slice, the user simply accesses the relevant records through their standard head slices: apparently the user accepts that the access structure between records and slices will guide the user to the relevant data. It is only when the user explicitly wants the selected attributes (and nothing else) that the user must refer to the new automatically generated slice. However, we acknowledge this possibility that the user wants to primarily access the volatile slices: therefore, we offer the additional record access structure between the volatile slices.

4.3 Multiple Relations

The third general case is that the attributes in <Attributes> belong to multiple relations. We feel that such a query expresses a relationship between records (relations). So, besides determining how records and slices are presented, we must determine how this relationship between records is presented. We propose that if there are more relations involved, each relation can be accessed through an indexed guided tour from the previous relation (in terms of the sequence in <Relations>). This means if <Relations> equals R1, R2, that from a record from R1 the associated records from R2 are accessible through an indexed guided tour.

As default representation for records and slices, we borrow their standard representation. Of course, we also want to express the information explicitly composed in the SELECT clause. We do so by adding to the records of the first relation R1 a new volatile slice with only the attributes in <Attributes>. This new slice obtains a default name, e.g. NEWSLICE, and it is connected through two links to the head slice of R1, e.g. HEAD-NEW and NEW-HEAD. Again we will also connect these new volatile slices using the default record access structure from the relation R1.

access structure within a record (between relations)	**entry slice**	**slices**	**slice access structure**	**presentation**	**record access structure**
indexed guided tour	head slice	as in data definition plus a volatile one for the first relation	for the first relation augmented with two new links	as in data definition	as in data definition of first relation

The principle underlying this part of the routine is that a user-specified association between records can best be represented by accessing the original relations via indexed guided tours. By offering a new automatically generated slice (page) with the explicitly selected attributes, the user can choose between this limited selection of information and a standard access structure offering a overview of the selected record associations.

Summarizing the above, the routine looks as follows:

1 relation	1 slice	selected slice as entry slice
1 relation	N slices	head slice as entry slice, one volatile slice in relation
N relations	M slices	indexed guided tour, one volatile slice for first relation

5. Adding Representation Information to Queries

The third step of our routine allows the user to explicitly override and adjust the default presentation (deduced in the first two steps) through "hints" in the query. While the first two steps use the principles of RMM, together with heuristics on the translation from SQL to RMDM, in the third step we allow the user to disagree with the proposed interpretation of the query semantics. It is important to note that in practice the user who composes queries may be different from the user (designer) responsible for the data definition.

Our third step assumes that we can use a preprocessor of the SQL queries that can interpret its input and produce (an SQL query and) user-specific presentation information. The purpose of this preprocessor is to combine the default representation from the data definition with the query dependent details to produce the presentation which the user likes best.

If we look at the first part of the routine (Section 4.1) we see that the user gets the slice from which he has selected some attributes. So, while the user may have specified a smaller number of attributes, he is presented with the entire slice. If the user explicitly wants this limited view of the slice, he should be able to specify that. For this reason we allow the user to write

```
SELECT NEW <Attributes>
```

in the query. The preprocessor interprets the NEW command in such a way that a new volatile slice is created with exactly the specified attributes. This new slice is accessible from the head slice, but it is presented as the entry slice.

The new volatile slice gets a default name, e.g. NEWSLICE. If the user explicitly wants to override that name, he can add the construct NAME <name> to the SELECT clause. Writing

```
SELECT <Attributes> NAME <name>
```

causes that <name> becomes the new name for the volatile slice.

Similarly, the user can override the default names for the links to and from the new slice. By adding the construct FORWARD LINK <name> and BACKWARD LINK <name>, the user can specify the new names for these links:

```
SELECT <Attributes> NAME <name> FORWARD LINK <link-name>
```

The last two parts of the routine (Section 4.2-3) show that in most cases a volatile slice is created. In a number of situations the user may not be interested in this slice, but is only interested in the selected associations between records. In that case the user may help the database system by notifying it that a new slice need not be created. By writing

```
SELECT NO NEW <Attributes>
```

in the query, the preprocessor is instructed to omit the creation of a new slice.

All three parts of the routine (Section 4.1-3) show that in general the default presentation from the data definition is used. The user may want to override that information in the query formulation.

- Writing INDEX or TOUR after a relation name in the FROM clause causes the database system to use an index or guided tour as the access structure for that relation. The next example causes the Person records to be accessible through an index:

```
FROM Person INDEX
```

- By writing INDEX or TOUR before a relation name (not the first) in the FROM clause, the database system will use the index or guided tour as the way to present the association between the relations. If for example the FROM clause looks like

  ```
  FROM Person, INDEX Employee
  ```

  then the association between Person and Employee records is given through an index (instead of the default indexed guided tour).

- The use of PAGE(x) after a relation name in the FROM clause, can override the default for the number of slices (records) shown on one page. In this case x is an integer (minimum 1), while we allow the user to write PAGE(*) to denote that all slices are shown on one page. The next expression results in one Hobby slice per page, if the attributes in the SELECT clause specify the Hobby slice:

  ```
  FROM Person PAGE(1)
  ```

Another aspect that can be influenced by the user is the implementation of the new record connecting access structure between the volatile slices. If the user is satisfied with the record access structure that only connects the head slices, the user can specify the construct HEAD ONLY in the FROM clause:

```
FROM HEAD ONLY <Relations>
```

Then only the head slices get connected (through the deafult access structure).

6. Conclusion and Future Work

In this paper we have described the outline of a routine to produce a hypermedia representation for volatile database output, specified by an SQL query. The routine uses the default representation specified in the data dictionary, but the user can override and adjust this default. It is important to realize that different people are involved. The designer of the system has the task to produce a suitable design of the query results that corresponds both to the semantics of the data queried and to the semantics of the query result itself. The designer must acknowledge that the person posing a query can override that design in order to efficiently use the query result in his work.

We are currently involved in a number of projects that require hypermedia (World Wide Web) presentation and navigation for information systems applications. One project concentrates on query interfaces for geographic information systems. A second project aims at providing travel, lodging and entertainment information for the disabled tourist in Europe. A third project aims at disclosing all museum information in the Netherlands through a Web server, including a general-purpose query interface for art- and history researchers. We intend to evaluate the proposed routine for generating hypermedia representations of database query results during the development phase of these projects.

References

[Díaz et al. 1995]

Díaz, A., Isakowitz, T., Maiorana, V. and Gilabert, G., *RMC: A Tool to Design WWW Applications.* Proceedings of the Fourth International World Wide Web Conference, pp. 559-566, Boston, 1995.

[Elmasri et al. 1990]

Elmasri, R. and Navate, S., *Fundamentals of Database Systems.* The Benjamin/Cummings Publishing Company, second edition, 1990.

[Garzotto et al. 1991]

Garzotto, F., Paolini, P. and Shwabe, D., *HDM - A Model for the Design of Hypertext applications.* Proceedings of the ACM Hypertext'91 Conference, ACM Press, pp. 313-328, 1991.

[Isakowitz et al. 1995]

Isakowitz, T., Stohr, E., and Balasubramanian, P., *RMM: A Methodology for Structured Hypermedia Design.* In Communications of the ACM 38, 8, pp. 34-44, (August) 1995.

[Pönighaus et al. 1996]

Pönighaus, R. and Mitlöhner, J., *Relational Databases and the World Wide Web: Automatic Generation of Hypertext based on Reverse-engineered Meta Information.* Proceedings of WebNet'96, the World Conference of the Web Society, pp. 401-406, San Francisco, 1996.

Teaching Biology on the Internet

Thomas S. Ingebritsen, Department of Zoology and Genetics and Project BIO, Iowa State University, USA,
tsingebr@iastate.edu

George G. Brown, Department of Zoology and Genetics and Project BIO, Iowa State University, USA,
g_brown@molebio.iastate.edu

John M. Pleasants, Department of Zoology and Genetics and Project BIO, Iowa State University, USA,
pleasant@iastate.edu

Abstract: Iowa State University, through a program called Project BIO, is using an innovative new approach to offer biology courses via the World Wide Web. The approach features on-line lectures similar to those a student might experience in a traditional classroom. Students listen to the lectures using RealAudio® while viewing lecture materials with a Web browser. The program, which began in Fall 1996 with two courses, has grown to eight courses for the 1997/98 academic year. The market for these courses includes: on-campus Iowa State University students, high school juniors and seniors, community college students, high school and community college biology teachers, and employees of life science companies.

Introduction

The internet is an exciting new medium for teaching biology at a distance. On-line biology courses offer a flexible learning environment where course materials can be accessed any time, day or night, from any location where a suitable computer connected to the internet is available. These courses support multiple learning styles (written, verbal and visual) and their content can be customized to meet the individual interests/needs of students. Another key feature of these courses is the ability to access authentic research databases as well as educational resources from other colleges and universities.

Iowa State University, through a program called Project BIO, is pioneering an innovative new approach for delivering courses and other educational material via the World Wide Web [1]. Iowa State is also exerting national leadership in the number and variety of biology courses that we are making available through this medium. Project BIO is a partnership involving educators in 7 departments and programs at Iowa State University, 14 of 16 Iowa community colleges and 43 Iowa high schools. The purpose of the partnership is to develop and share biology education resources via the internet. For more information see the Project BIO World Wide Web site (http://project.bio.iastate.edu).

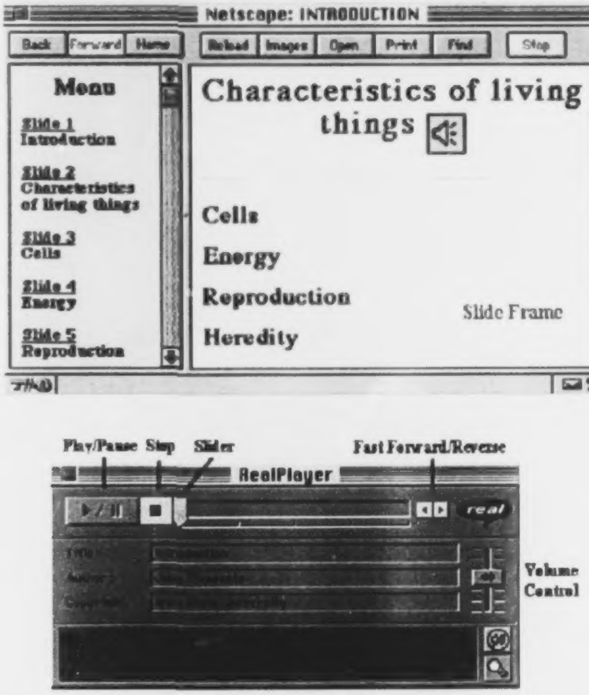
Approach and Technology

Our approach features on-line lectures that are similar to presentations made in a traditional on-campus classroom. These presentations, which are available 24 hrs/day via the internet, consist of a set of slides that are accessed as Web pages and an audio explanation of the material on the slides. The audio portion of the presentation is being delivered using a new audio streaming technology called RealAudio®. Our approach represents a significant advance over the typical internet approach of delivering educational information using text and static images.

---

[1] Funding for this project was provided by grants from the Kellogg Foundation Vision 2020 project and the Howard Hughes Medical Institute. The following Iowa State University administrative units also provided support for this project: Provost , College of Agriculture , College of Liberal Arts and Sciences , Botany Department , Office of Biotechnology , Zoology & Genetics Department.

Fig. 1 (Top Panel) shows a typical lecture window consisting of a menu frame and a frame for displaying slides. Slides can be accessed sequentially or randomly using the menu. The RealPlayer® functions as a helper application that is linked to the World Wide Web browser. [Fig. 1] (Bottom Panel) shows the RealPlayer® control panel. The audio portion of the lecture is accessed by clicking on an audio button which is present on each slide.


Netscape: INTRODUCTION
Back
Forward
Home
Reload
Images
Open
Print
Find
Stop
Menu
Slide 1
Introduction
Slide 2
Characteristics of living things
Slide 3
Cells
Slide 4
Energy
Slide 5
Reproduction
Slide Frame
Play/Pause
Stop
Slider
Fast Forward/Reverse
RealPlayer
real
Volume Control

Characteristics of living things

Cells

Energy

Reproduction

Heredity

Figure 1: Typical lecture window (top) and RealPlayer® (bottom)

Courses and Audience

Iowa State University began offering on-line biology courses during Fall semester 1996 with two courses, "Biotechnology in Agriculture, Food and Human Health" and "Introduction to Basic Microbiology" [Tab. 1). These were the first two courses from ISU to be taught exclusively via the World Wide Web. During Spring semester 1997 four courses were offered with a total enrollment of 145 students. This represented a 5-fold increase over the Fall 1996 enrollment. About 60% of the students are off campus students and 40% are on campus. We expect to offer 8 on-line biology courses during the1997/98 academic year [Tab. 2]. Most of the courses have been adapted from existing on-campus courses and involve faculty who also regularly teach the courses to on-campus students in a traditional classroom setting. All of these courses are accessed from the Project BIO World Wide Web site (http://project.bio.iastate.edu). Byway of comparison CASO's: The Internet University (http://www.caso.com/), an authoritative guide to on-line courses, lists only four other universities that offer a total of five on-line biology courses.

Target audiences for our on-line biology classes include: on-campus Iowa State University students, high school juniors and seniors, community college students, high school and community college biology teachers, and employees of life science companies [Tab. 2].

BEST COPY AVAILABLE

Semester	Course Number	Course Title	On Campus	Off Campus	Total
Fall 96	MIPM 302	Introduction to Basic Microbiology	7	3	10
	Gen 308/508	Biotechnology in Agriculture, Food and Human Health	2	17	19
		Totals	**9**	**20**	**29**
Spring 97	Biol 109	Introductory Biology, Non-majors	7	29	36
	Biol 201	Principles of Biology, Majors	6	37	43
	MIPM 302	Introduction to Basic Microbiology	36	7	43
	Gen 308/508	Biotechnology in Agriculture, Food and Human Health	11	12	23
		Totals	**60**	**85**	**145**

Table 1: Project BIO On-line Biology Courses, Fall 96 and Spring 97 Enrollments

Course Number	Course Title	Target Audience
Biol 109	Introductory Biology, Non-majors	ISU students, high school students
Biol 201	Principles of Biology, 1st Semester, Majors	ISU students, high school students
Biol 202	Principles of Biology, 2nd Semester, Majors	ISU students, high school students
MIPM 201	General Microbiology, Non-majors	ISU students, high school students
Zool 155	Basic Human Physiology and Anatomy, Non-majors	ISU students, high school students
MIPM 302	Introduction to Basic Microbiology	ISU students, community college students, industry employees
Gen 308/508	Biotechnology in Agriculture, Food and Human Health	ISU students, biology teachers, community college students, industry employees
MIPM 501X	Advanced Microbiology	ISU students, biology teachers, industry

Table 2: Project BIO On-line Biology Courses, Academic Year 1997/98

High School Students

Iowa's Post-secondary Enrollment Options Act [1] allows 11th and 12th grade students to enroll part time at an eligible community college, state university, or private college or university. The student's high school or school district pays for the cost of tuition, textbooks, materials and fees up to $250. Students earn both high school and college credits for the courses taken. This program provides opportunities for high school juniors and seniors to get a head start on college. It also makes challenging courses available for talented and gifted students. This program is especially important for small rural school districts in Iowa that often do not have the resources to offer Advanced Placement (AP) courses. For example, only 123 of ~400 high schools in Iowa offer an advanced placement biology course [2]. Moreover, to our knowledge there are no high schools in Iowa that offer college level biology courses for students heading for non-science majors.

---

[1] Excerpts from Iowa Post-secondary Enrollment Options Act (Chapter 261C) (http://project.bio.iastate.edu/Courses/pseoacod.htm)

[2] Educational Testing Service, personal communication.

Students who wish to exercise the Post-secondary Enrollment Option face two key problems in taking these courses. One is distance from the community college or university offering the courses. This issue is particularly important in Iowa where a significant proportion of the population lives in rural communities. A second important issue is scheduling. College classes are generally longer than high school classes, there is often significant travel time to a site where college classes are offered, and it is often difficult for high school students to schedule college classes around extracurricular activities.

Our on-line biology courses offer an ideal solution to many of these problems. Students can listen to on-line lectures at home or at their school at a time that is convenient to them. They can listen to an on-line lecture over two or more class periods.

Biology Teachers

In order to maintain their teacher certification high school biology teachers in Iowa need to take 6 college credits every 5 years and community college biology teachers need to take 4 credits during this same time period. Typically teachers must take courses or workshops in evenings or during the summer to fulfill these certification requirements. Evening courses are difficult to fit into a busy schedule while taking courses in the summer means loss of income. In addition the choice of courses in the evenings or during the summer semester is very limited. On-line courses offer an attractive alternative because teachers can work on these courses during the evenings or weekends at a time that is convenient for them.

On-Campus Students

About 40% of the students enrolled in our on-line biology classes are on-campus ISU students. Lower division biology courses are particularly attractive because of the very large class size (200-500 students) of traditional on-campus sections of these courses. They are typically heavily subscribed or over-subscribed making it difficult for students to fit the courses into their schedules. By way of contrast, students can access on-line lectures from their dorm rooms or from computer labs (on-campus or in the dorms) at anytime, 24 hours a day. The on-line biology courses are smaller and offer a more intimate learning environment with greater access to the instructor through the use of technologies such as e-mail, interactive Web pages and chat.

Student Access

Technology

Students need a computer (PC or a Mac) that is connected to the Internet at a minimum speed of at least 14.4 kbps. Two pieces of client software are required to access the lectures. A World Wide Web browser with frames capability for accessing the slides and the Real Player® for accessing the audio component of the lectures. Both software items can be downloaded from the internet [1] and are available at no charge to students.

On-Campus Students

They can access the courses using their own computer in their home, apartment or dormitory room. They can also use one of many computer labs located in their dormitories or on the campus. All dormitory rooms have ethernet ports providing students with fast access to the internet.

---

[1] Two of the most popular browsers are Netscape Navigator (http://home.netscape.com/) and Microsoft's Internet Explorer (http://www.microsoft.com/). The RealPlayer® can be downloaded from (http://www.real.com/).

Off-Campus Students

Some of these students are able to access the course using a computer at home or work with access to the internet. However a major problem here is that more than 3/4 of the potential audience for internet courses does not have internet access [1]. A major factor is cost of the technology and this clearly discriminates against economically disadvantaged students. We are attempting to deal with this problem by working with Iowa high schools and community colleges to help them set up public access terminals that their students and faculty can use to access our on-line biology courses. To date these terminals have been established in 43 high schools and on five community college campuses.

Authentic Learning Experiences

An exciting educational feature of the internet is that it is possible for students to access and utilize information in authentic research databases as part of lectures or learning activities. In the "Biotechnology in Agriculture, Food and Human Health" course we have exploited this possibility in several learning activities that have been developed for the course. For example in an assignment called "Genetic Diseases", students are required to write a report about a genetic disease of their choice based on information obtained from the On-line Mendelian Inheritance in Man (OMIM) database (http://www3.ncbi.nlm.nih.gov/omim/).

Four of the learning activities in the biotechnology course are on-line lab simulations. In one activity called "Cloning by Computer", students access DNA sequences in the GenBank database (http://www.ncbi.nlm.nih.gov/) and then use a word processing program to cut and paste the DNA sequences together. This simulates a key step in the genetic engineering process. Two other activities involve taking the students through photographs of a wet laboratory demonstration and then having the students interpret data obtained in the lab. The third, called the Virtual FlyLab is a full-featured lab simulation developed at California State University, Los Angeles. In this simulation, students are able to design and interpret the results of virtual fly matings conducted on-line. The URL for the Virtual FlyLab is http://vflylab.calstatela.edu/edesktop/VirtApps/VflyLab/IntroVflyLab.html.

All of the learning activities developed for the biotechnology course are in the public domain and can be accessed from the biotechnology course homepage (http://project.bio.iastate.edu/courses/gen308/Home/Homepage1SS.html).

Course Administration

This is handled by software called ClassNet (http://classnet.cc.iastate.edu/) that was developed at Iowa State University. This software allows for on-line testing through an interactive page that can be accessed via a Web browser. Tests can include multiple choice, fill-in-the-blank and essay questions. The multiple-choice and fill-in-the-blank questions are machine graded whereas the instructor manually grades the essay questions. Students are required to identify a proctor in their local community who verifies the identity of the student and supervises the test. The proctor must supply a password before a student can access a test.

Classnet also provides several mechanisms for student/student and student/instructor interaction. These include the ability to send e-mail messages to students or instructors from the Web browser, a threaded class discussion forum in which messages can be sent using an interactive Web page and a chat feature that provides for real time communication.

---

[1] CommerceNet/Nielsen Internet Demographics(http://www.commerce.net/work/pilot/nielsen_96)

Course Development

Project BIO Resource Center

The purpose of the center is to assist faculty in the development of on-line courses by providing technology resources, technical assistance and training. Technology in the center includes: 1) a World Wide Web server, 2) an 80 stream RealAudio/Video® server, 3) a sound-proof room with facilities for recording, digitizing and editing audio files, and 4) four general purpose World Wide Web authoring computers (3 Macintosh Power PC's and one Hewlett Packard VectraVL4). The facility also has a sound proof room which we will use to develop video content for the on-line courses. The room contains equipment for digitizing and editing video files and for recording voice-overs. Video content will be delivered using RealVideo® technology. The staff of the resource center includes the Professor-in-Charge, Tom Ingebritsen, a Technology Specialist, a Secretary, a Graduate Assistant and 3 Undergraduate Assistants. The staff maintains the technology, assists with the preparation and editing of course materials and provides training for faculty and staff. Plans for the Resource Center are shown in [Fig. 2]

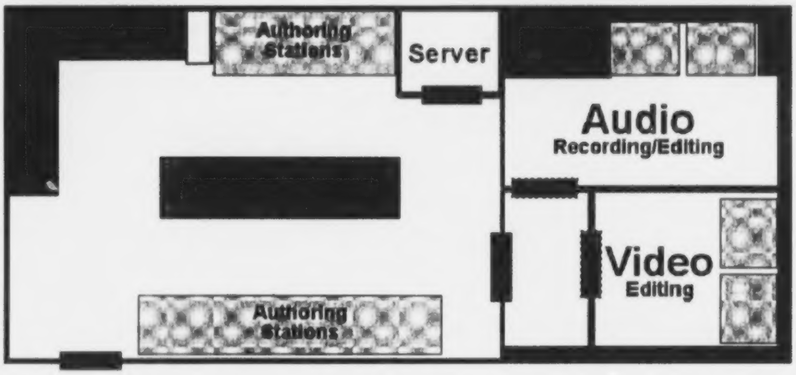


Figure 2: Plans for Project BIO Resource Center

Web pages

The majority of the Web authoring was done using a what-you-see-is what-you-get HTML authoring program called Claris Homepage (http://www.claris.com). In some cases the HTML editing capabilities of Netscape Navigator, Internet Explorer and Microsoft Office were also used.

Images and diagrams used in the lectures were created and/or edited using Adobe Photoshop (http://www.adobe.com) and Macromedia Freehand (http://www.macromedia.com). Some textbook images were used with appropriate attribution and permission from the publishers.

Audio

The audio part of the lectures were either directly recorded on a Macintosh Power PC computer (7500/100) computer or recorded using an analog tape recorder. In the latter case the contents of the analog tape were then input into the computer and digitized. SoundEdit 16 software was used to digitize and edit the audio files. The audio files were compressed and converted to the RealAudio format using RealAudio® Encoder software which is available from Progressive Networks (http://www.real.com) at no charge. Audio files were typically compressed from ~150 Mb for an uncompressed 1 hour lecture to 4-5 Mb in the RealAudio® format.

BEST COPY AVAILABLE

Information Access in the Web

Luca Iocchi, Daniele Nardi
Dipartimento di Informatica e Sistemistica
Università di Roma "La Sapienza"
via Salaria 113, 00198 Roma, Italy
E-mail: `{iocchi,nardi}@dis.uniroma1.it`

Abstract: While the amount of information available on line in the Internet or in the World Wide Web is increasingly growing, many different systems have been developed to help the user to locate or gather information of interest. In this paper we discuss the main approaches to the design of such systems focusing in particular on the methods used to represent the information available in the Web.

1 Introduction

The management of large amounts of information is a critical task for an organization or a workgroup. The growth of information available in various repositories within an organization or on-line through the World Wide Web, makes it difficult for users to gather information of interest, while browsing and keyword-based search become more and more ineffective.

The World Wide Web can be seen as a huge collection of different information sources, that are on-line databases, information systems or simply HTML pages, containing a very large body of information from very different areas. An individual user is usually interested only in a small portion of this information, so s/he needs tools for an effective organization of her/his own information and for an intelligent access to information sources in order to retrieve important or relevant information in an effective and efficient way.

Recent research work is focused on the design of systems that help the user to locate or gather desired information in a world with many different and heterogeneous information sources. In this paper we discuss the main design elements of these systems, especially outlining the different approaches arising from the choice of different methodologies for representing information.

2 Surfers vs Hunters

A first distinction among systems is based on their main goal. We identify a first group in which we consider systems whose main goal is to learn a user profile to assist the navigation of users through the Web (we refer to such systems as *surfers*). The second group is characterized by the integration of various information systems and a query answering mechanism (we call them *hunters*).

Two examples of surfers are WebWatcher [Armstrong et al., 1995] and Letizia [Lieberman, 1995]. They present a Web-based interface to interactively assist user browsing, learning what s/he usually looks for (using several navigation tasks as training set) and trying to anticipate which links will be followed. Surfers do not take control over the user, they only suggest possibly relevant links to follow.

They differ in the function to be learned. WebWatcher learns the probability that an arbitrary user will choose a link starting from a certain page to achieve a goal, which corresponds the function

$$UserChoice? : Page \times Goal \times Link \rightarrow [0, 1]$$

while Letizia tries to infer user goals from his browsing behavior. Pages, links and goals are represented by lists of keywords or by feature vectors, each feature indicating the occurrence of a particular word within the text. A comparison between different machine learning techniques to be used in this task is presented in [Armstrong et al., 1995]. A related approach is followed in [Cohen and Singer, 1996], in which the system learns to query the Web, that is it learns how to use a search tool (which keywords have to be used) to retrieve important information for the user.

We do not consider systems whose main goal is an automatic classification of documents. They usually represent documents as a feature vector with each component corresponding to the frequency of a word in the document or the salience in the text of a subject category, constructing personalized information filters. Such a representation is used, however, for special class of textual documents such as e-mail, Usenet news and Web pages and machine learning techniques have been proposed to learn rules that classify them (see for example [Cohen, 1996, Bloedorn et al., 1996, Goan et al., 1996]).

Hunters differ from surfers in building a (virtual) common model of the relevant information space. They act as spiders through the Web gathering information for the user. Many special purpose agents (we call them *information brokers*) have been developed to retrieve a particular type of information from predefined information sources. For example one can use BargainFinder [Krulwich, 1996] to find where to buy the last CD of his favorite artist for the best price, or ContactFinder [Krulwich and Burkey, 1996] to identify a person who can help him in solving a problem. In the following we focus on Global Information Management Systems, that are general purpose systems in which there is an explicit representation of domain and information sources.

3 Global Information Management Systems

Global Information Managements Systems (GIMSs) provide a framework to integrate different and heterogeneous information sources into a common domain model. An information source can be an on-line database accessible through the Web or a simple HTML page or a plain text file. Information units are individual elements of information coming from information sources. The user interacts with the GIMS as a single information system, so that s/he can ignore data models used in the individual sources, and accesses information through query-answering mechanisms.

A basic distinction among GIMSs can be done considering different methods to represent information. First we address feature-based representation (or keyword representation), in which documents are represented with feature vectors (or simply list of keywords), like in the well known keyword-based search engines. Second we explore the work from the Database community, in which both systems using a conceptual data model to represent information domain and Web Query Languages are developed. Finally, systems using a Knowledge Representation approach are presented. They use an explicit representation of knowledge about domain and information sources and automatic reasoning tools to answer user queries.

3.1 Feature-based Representation of Documents

The representation of text documents through a feature vector is very simple but also useful. A vector whose features are specific words describing document contents is associated to each document. A particular case is that in which documents are represented by a list of keywords.

In this class we include the well known keyword-based search engines, such as Altavista, Lycos, Yahoo!, and many others. They are equipped with soft-bots that explore the entire Web reading documents and indexing them according to some key. Then they allow for retrieving the previously analyzed documents from specified keywords. We must observe that these systems provide a limited integration of information sources, as they typically consider only HTML or plain text documents, nonetheless they are broadly used being easily accessible for the user. We do not list specific features of these systems, as one can find a lot of surveys and comparisons on-line in the Web. We only point out the importance of organizing Web documents into a hierarchical structure, as in Yahoo!, even though in this system there are different and not uniform subdivision criteria within a single hierarchy (e.g. is-a and part-of relationships, geographic and time divisions, etc.).

3.2 Database Approaches

In the Database community we find two kinds of proposals: systems to integrate different information sources and declarative languages to query the Web. We do not specifically address tools for database integration or federated databases, since they rely on the presence of a schema describing sources and in highly structured data, while Web documents are usually unstructured or semi-structured.

A notable example of GIMS using database technology to represent information is Tsimmis [Chawathe et al., 1994], which describes the common model with the OEM (Object Exchange Model) language and the associated query language, OEM-QL, is an SQL-like language. It makes use of ***translators*** to translate data object into a common information model and queries into requests for an information source, while ***mediators*** embed the knowledge necessary for processing a specific type of information, knowing the contents of information sources. This distinction between translators and mediators allows different mediators to work independently. Each mediator needs to know which sources it will use to retrieve information. In this way it is possible to work without a global database schema. Furthermore, mediators and translators can be automatically generated from high level descriptions of the information processing they have to accomplish. Constraint Manager units are also used to define integrity constraints which specify semantic consistency requirements, while classifiers and extractors can be used to extract information from unstructured documents (e.g. plain text files, mail messages, etc.) into the domain model. The Classifier/Extractor components of Tsimmis are used in order to extract information from unstructured documents.

Another proposal along these lines is constituted by the ARANEUS Project [Atzeni et al., 1997], whose aim is to make explicit the schema according to which the data are organized in so-called structured servers, and then use this schema to pose queries in a high level language instead of browsing the data. Even though the ability to construct structured description of the information in the Web enables the system to answer user queries, the approach has the following drawbacks that are typical of a Database perspective: 1) Araneus works only on a particular kind of Web sites and pages, which have a clearly specified structure, not on generic ones; 2) the user has to completely specify the relational schema corresponding to the site data; there is no automatic translation from the site to the database; 3) there is no hint for automatic search and conceptualization of WWW sites similar to prototypical ones indicated by the user.

WAG (Web At a Glance) [Catarci et al., 1997] is a system that assists the user in the construction of a conceptual view of Web pages relevant to her/his own interests. The main difference with other database approaches is that, instead of requiring an explicit description of the sources, WAG attempts to semi-automatically classify the information gathered from various sites based on the conceptual model of the domain of interest. The result of such a classification is fully materialized. In addition, WAG provides a visual interface to query the databases (each one related with a specific domain or sub-domain) resulting from the integration of the information extracted from the various sites.

A second research area in the Database community involves the development of declarative languages to query the Web. Web Query Languages proposed in literature are W3QL [Konopnicki and Shmueli, 1995], WebLog [Lakshmanan et al., 1996], WebSQL [Mendelzon et al., 1996]. Conceptual models of the World Wide Web are presented to specify semantics of these languages. In particular in [Mendelzon et al., 1996] a "virtual graph" is used to represent the hypertextual documents in the Web. Systems using a Web Query Language do not maintain a global model of an application domain, instead they allow the user to interact with Web search engines or indexes built from robots in a transparent way. Many of the problems one encounters using indexes, such as information updates or the lack of representation of the structures in the documents, are not addressed in these systems. However the possibility of capturing the structure of a hypermedia network, explicitly describing links between documents, and the introduction of the "query locality" concept to measure the cost to answer a query are important elements in the development of effective and efficient systems.

4 Knowledge-based GIMSs

Knowledge-based GIMSs are systems using a Knowledge Representation (KR) approach for information sources representation, data acquisition and query processing. Many logical frameworks are used to represent information and many Knowledge Representation systems are used to reason about them.

The main design element for these systems is the KR language. While the most important tasks they have to accomplish are automatic information acquisition, that is useful to build and maintain knowledge bases, as well as query answering using query-planning techniques.

System	KR Language	Vocabulary Problem
Information Manifold [Levy et al., 1996]	CARIN	Unique vocabulary
SIMS [Arens et al., 1996]	LOOM	Manual mapping
Internet Softbot [Etzioni and Weld, 1994]	UWL	Unique vocabulary
OBSERVER [Mena et al., 1996]	CLASSIC	Semi-automatic mapping
Information broker [Fikes et al., 1996]	Context logic	Automatic mapping
Infomaster [Geddis et al., 1996]	Datalog-like	Unique vocabulary

Table 1: Knowledge Representation in GIMSs

4.1 Knowledge Representation

A GIMS represents both the application domain and contents of information sources, using usually a single KR language. [Tab. 1] shows different logical frameworks used by some of the implemented systems. As the knowledge base of a GIMS is formed by a collection of concepts related by semantical and hierarchical relationships, it seems that formalisms able to represent taxonomic knowledge, such as Description Logics, are valuable in this context for their capability to represent hierarchical concept structures.

One critical problem arising from the integration of different descriptions of information sources is the vocabulary problem. It is due to the presence of possibly different terms representing the same concept in the description of a source or an information unit. There are three possibilities to face this problem: (i) unique vocabulary, that is forcing the description of information sources and domain model to share the same vocabulary; (ii) a manual mapping, that is relationships between similar concepts are hand-coded; (iii) automatic (or semi-automatic) mapping, in which the system takes advantage of existing ontology systems (Ontolingua, WordNet) that provide synonym, hypernym and hyponym relationships between terms. In [Tab. 1] we also show how systems address the vocabulary problem. In particular OBSERVER [Mena et al., 1996], which is based on the interaction of different ontologies, addresses the problem both in the definition of the ontologies and by providing a tool for defining semantical relationships among terms of different ontologies. While in the information broker presented in [Fikes et al., 1996, Farquhar et al., 1995] the use of linguistic tools provided by Ontolingua is proposed.

Let us notice that using a unique vocabulary can lead to an extremely rigid system. On the other hand using linguistic tools is very powerful to solve questions about the terminology and to retrieve information, even though it involves information loss due to the use of terms not completely suitable to describe information units.

4.2 Information Acquisition

An important feature for a GIMS is the possibility of identifying interesting information sources unknown to the user and to automatically gather from them relevant information units. In other words, tools to scale up with the growth of the information space are needed. The discovery of new information sources, the extraction of information units within them and the interpretation of data coming from these sources are all problems related to information acquisition.

This issue is rarely addressed in most systems, as they force the user to hand-code information sources' models. The main exceptions are ShopBot and ILA [Perkowitz et al., 1996]. ShopBot addresses the extraction problem learning how to access an on-line catalog (via an HTML form) and how to extract information about products. It uses an unsupervised learning algorithm with a small training set. Whereas ILA (Internet Learning Agent) is focused on the interpretation problem. It learns how to translate information source's output into the domain model, using a set of descriptions of objects in the world.

4.3 Query Processing

A significant body of work on agents able to reason and make plans for query answering has been developed. The use of planning techniques to retrieve information requested by a user query has been very common in this context.

In Information Manifold [Levy et al., 1996] the contents of information sources are described by query expressions that are used to determine precisely which sources are needed to answer the query. The planning algorithm first computes information sources relevant to each subgoal, next conjunctive plans are constructed so that the soundness and completeness of information retrieval and the minimization of the number of information sources to be accessed are guaranteed. In this system, interleaving planning and execution is a useful way to reduce the cost of the query during plan execution.

The Infomaster [Duschka and Genesereth, 1996] planning method is similar to the Information Manifold one. Infomaster guarantees a semantically correct and source-complete plan generation in a very expressive representation language, but clearly separates query planning and plan execution.

SIMS [Arens et al., 1996] defines operators for query reformulation and uses them to select relevant sources and to integrate available information to satisfy the query. The system applies these operators first to reformulate the query according to the information sources model, and then to identify the information sources to access. Since source selection is integrated into the planning system, SIMS can use information about resource availability and access costs to minimize the overall cost of a query.

Previously described systems rely on a closed world assumption, that is they assume that domain model contains all information needed and that all unavailable information does not exist. On the contrary Internet Softbot [Etzioni and Weld, 1994] provides a framework to reason with incomplete information [Etzioni et al., 1992, Etzioni et al., 1994], executing sensing actions to provide forms of local closure, in other words to verify the actual presence of information in the source during plan execution.

5 Conclusion

We conclude by stating that an intelligent access to information in the Web deeply relates to the representation of information. In fact simple representation languages, used in most search tools, have severe limitations for an effective retrieval of information. On the other hand, more structured representations of knowledge are difficult to build automatically, but can provide more effective tools for information access.

References

[Arens et al., 1996] Arens, Y., Knoblock, C. A., and Shen, W. (1996). Query reformulation for dynamic information integration. *Journal of Intelligent Information Systems.*

[Armstrong et al., 1995] Armstrong, R., Dayne, F., Joachims, T., and Mitchell, T. (1995). WebWatcher: A learning apprentice for the World Wide Web. In *Proc. of AAAI Spring Symposium on Information Gathering form Heterogeneous Distributed Environments.*

[Atzeni et al., 1997] Atzeni, P., Mecca, G., and Merialdo, P. (1997). Design and maintenance of data-intensive web sites. Technical Report 25, Dip. di Informatica e Automazione, Università di Roma Tre.

[Bloedorn et al., 1996] Bloedorn, E., Mani, I., and MacMillan, T. R. (1996). Representational issues in Machine Learning of user profiles. In *Proc. of AAAI Spring Symposium on Machine Learning in Information Access.*

[Catarci et al., 1997] Catarci, T., Chang, S. K., Nardi, D., and Santucci, G. (1997). Wag: Web-at-a-glance. Technical Report 03-97, Dip. di Informatica e Sistemistica, Università "La Sapienza" di Roma.

[Chawathe et al., 1994] Chawathe, S., Garcia-Molina, H., Hammer, J., Ireland, K., Papakonstantinou, Y., Ullman, J., and Widom, J. (1994). The TSIMMIS Project: Integration of Heterogeneous Information Sources. In *Proc. of IPSJ Conference*, pages 7–18.

[Cohen, 1996] Cohen, W. W. (1996). Learning Rules that Classify E-Mail. In *Proc. of AAAI Spring Symposium on Machine Learning in Information Access.*

[Cohen and Singer, 1996] Cohen, W. W. and Singer, Y. (1996). Learning to query the web. In *AAAI Workshop on Internet-Based Information Systems.*

[Duschka and Genesereth, 1996] Duschka, O. M. and Genesereth, M. R. (1996). Query planning in Infomaster. `http://logic.stanford.edu/people/duschka/papers/Infomaster.ps`.

[Etzioni et al., 1994] Etzioni, O., Golden, K., and Weld, D. (1994). Tractable Closed World Reasoning with Updates. *Proceedings of the Fourth International Conference on the Principles of Knowledge Representation and Reasoning (KR-94).*

[Etzioni et al., 1992] Etzioni, O., Hanks, S., Weld, D., Draper, D., Lesh, N., and Williamson, M. (1992). An approach to planning with incomplete information. *Proceedings of the Third International Conference on the Principles of Knowledge Representation and Reasoning (KR-92).*

[Etzioni and Weld, 1994] Etzioni, O. and Weld, D. (1994). A Softbot-Based Interface to the Internet. *CACM*, 37(7).

[Farquhar et al., 1995] Farquhar, A., Dappert, A., Fikes, R., and Pratt, W. (1995). Integrating information sources using context logic. In *Proc. of AAAI Spring Symposium on Information Gathering form Heterogeneous Distributed Environments.*

[Fikes et al., 1996] Fikes, R., Farquhar, A., and Pratt, W. (1996). Information brokers: Gathering information from heterogeneous information sources. In *Proc. of the Ninth Florida Artificial Intelligence Research Symposium (FLAIRS'96).*

[Geddis et al., 1996] Geddis, D. F., Genesereth, M. R., and Keller, A. M. (1996). Infomaster and Information Integration. In *Proceedings of WWW5 Workshop on Artificial Intelligence-based tools to help W3 users.*

[Goan et al., 1996] Goan, T., Benson, N., and Etzioni, O. (1996). A grammar inference algorithm for the World Wide Web. In *Proc. of AAAI Spring Symposium on Machine Learning in Information Access.*

[Konopnicki and Shmueli, 1995] Konopnicki, D. and Shmueli, O. (1995). W3QS: A query system for the World Wide Web. In *Proceedings of the Twentyfirst International Conference on Very Large Data Bases (VLDB-95)*, pages 54–65.

[Krulwich, 1996] Krulwich, B. (1996). BargainFinder. `http://bf.cstar.ac.com/bf/`.

[Krulwich and Burkey, 1996] Krulwich, B. and Burkey, C. (1996). The ContactFinder agent: Answering bulletin board questions with referrals. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI-96).*

[Lakshmanan et al., 1996] Lakshmanan, L., Sadri, F., and Subramanian, I. N. (1996). A declarative language for querying and resctructuring the Web. In *Proc. of 6th International Workshop on Research Issue in Data Engineering (RIDE-96).*

[Levy et al., 1996] Levy, A. Y., Rajaraman, A., and Ordille, J. J. (1996). Querying-Answering Algorithms for Information Agents. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI-96).*

[Lieberman, 1995] Lieberman, H. (1995). Letizia: an Agent that assist Web Browsing. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95).*

[Mena et al., 1996] Mena, E., Kashyap, V., Sheth, A., and Illarramendi, A. (1996). OBSERVER: An approach for query processing in global information systems based on interoperation across pre-existing ontologies. In *Proceedings of the Conference on Cooperative Information Systems.*

[Mendelzon et al., 1996] Mendelzon, A., Mihaila, G. A., and Milo, T. (1996). Querying the World Wide Web. In *Proc. of PDIS'96.*

[Perkowitz et al., 1996] Perkowitz, M., Doorebons, R. B., Etzioni, O., and Weld, D. S. (1996). Learning to understand information on the Internet: an example-based approach. *Journal of Intelligent Information Systems.*

'Interactivity' – Tracking a New Concept

Jens F. Jensen, Department of Communication, Aalborg University, Denmark, jensf@hum.auc.dk

Abstract: The purpose of this paper is to track the concept of 'interactivity'. The paper starts with a short discussion of the concept's current placement in the fields of media and communication studies and its background in other traditions. This is followed by the presentation of various representative attempts at definitions from academic studies. Finally, based on this presentation, a new definition of 'interactivity' is suggested.

»in´ter.ac´tive
1. new technology that will change the way you *shop*, *play* and *learn*
2. a zillion-dollar *industry* (maybe)«.

The above quote is a quick, dictionary-like keyword definition of the concept 'interactive' as it appeared on the cover of *Newsweek* on May 31, 1993. The quote is in many ways characteristic. In recent years, expectations of 'interactivity' and new 'interactive media' have been pushed to the breaking point in terms of what will become technologically possible, in terms of services that will be offered, in terms of economic gain, etc. Along with terms like 'multimedia', 'hypermedia', 'convergence' and 'information superhighway', 'interactivity' is presumably among the words currently surrounded by the greatest amount of *hype*. The concept seems loaded with positive connotations along the lines of high tech, hypermodernity and futurism, along the lines of individual freedom of choice, personal development, self determination,–and even along the lines of folksy popularization, grassroots democracy, and political independence. At the same time, it seems relatively unclear just what 'interactivity' means. The positiveness surrounding the concept and the frequency of its use seem, in a way, to be reversely proportional to its precision and actual content of meaning. 'Interactivity' is currently one of the media community's most used *buzzwords*. Maybe this isn't so surprising after all. The meaning of professional terms–including scientific and academic terms–is often watered down once they win popular acceptance in daily usage. And with the explosive growth and decided success of interactive technologies in recent years in the form of computers, multimedia, Internet, WWW, etc.–where it can be said that culture has lived out what we might call 'the interactive turn'–'interactivity' has naturally entered common usage. This kind of confusion of concepts is, however, inappropriate in an academic situation where it is necessary to know relatively precisely what terms refer to and which differences they make. At the same time, the concept of 'interactivity' has a longer and more complicated tradition behind it than first meets the eye. There are, therefore, many good reasons to leave the *hype* and *buzz* behind and take a closer look instead at the background and construction of the concept of 'interactivity'.

'Interactivity'–Media Studies' Blind Spot?

While *Newsweek*, as previously cited, dared to publish a cover with a refreshing keyword definition, more serious definitions are harder to find in common reference works and handbooks from the fields of media and communication. Here the term 'interactivity' is most notable for its absence. Naturally, this blind spot has an explanation. One way to clarify what may be blocking the view–and at the same time establish a framework for understanding the various concepts of interactivity currently in circulation–is to use the media typology developed by [Bordewijk & Kaam 86]. Their typology is based on two central aspects of all information traffic: the question of who owns and provides the information, and who controls its distribution in terms of timing and subject matter. By cross-tabulating these two aspects in relation to whether they are controlled by either a centralized information provider or a decentralized information consumer, a matrix appears with four principally different communication patterns, as illustrated in [Fig. 1] [see also Jensen 96a] [& Jensen 96b]

	Information produced by a central provider	*Information produced by the consumer*
Distribution controlled by a central provider	**1) TRANSMISSION**	**4) REGISTRATION**
Distribution controlled by the consumer	**3) CONSULTATION**	**2) CONVERSATION**

Figure 1: Bordewijk & Kaam's matrix for the four communication patterns

1) If information is produced and owned by a central information provider and this center also controls the distribution of information, we have a communication pattern of the *transmission* type. This is a case of one way communication, where the significant consumer activity is pure reception. Examples would be classic broadcast media such as radio and TV (but also, e.g., live broadcasts of conferences, TV, or multimedia via the MBone). 2) If the exact opposite occurs and information is produced and owned by the information consumers who also control distribution, we have a *conversation* pattern of communication. This is a case of traditional two way communication, where the significant consumer activity is the production of messages and delivery of input in a dialog structure. Typical examples would be the telephone, e-mail, newsgroups, IRC, etc. 3) If information is produced and owned by an information provider, but the consumer retains control over what information is distributed and when, it is a *consultation* communication pattern. In this case, the consumer makes a request to the information providing center for specific information to be delivered. Here the characteristic consumer activity is one of active selection from available possibilities. Typical examples would be various *on-demand* services or on-line information resources such as FTP, Gopher, WWW etc. 4) Finally, if information is produced by the information consumer, but processed and controlled by the information providing center, we have a *registration* communication pattern. In this communication pattern the center collects information from or about the user. In this case, the characteristic aspect is the media system's storage, processing, and use of the data from or about the user. Typical examples would be various types of surveillance, registration, and logging of computer systems.

Among these four information patterns, *transmission* is the only one that is characterized by one way communication. In other words, there is no back-channel that makes an information flow possible *from* the information consumer *to* the media system. Until now, communication and media studies has primarily based its models and insights on the transmission pattern because of the dominant role played by mass communication research. Current media developments including the arrival of 'new media' such as the Internet, intranets, networked multimedia, WWW, etc. have been more or less singularly characterized by a movement away from the transmission pattern toward the other three media patterns. These new media, which open up the possibility for various forms of input and information-flow from information consumers to the system, can hardly be described using traditional one way models and terminology. Seen from this perspective, it might well be claimed that as developments proceed, existing media theory is increasingly less able to explain current media phenomena. Perhaps for these reasons, among others, the established communication research community has developed blind spots in relation to new interactive media. This general problem can only be mentioned briefly here, as we proceed to follow another, more specific trail ...

'Interactivity'–The Background Behind the Concept

As [Jäckel 95], among others, has pointed out, the concept 'interactivity' extends–perhaps not surprisingly–from the concept of 'interaction'. A concept which generally means: 'exchange', 'interplay', 'mutual influence'. However, if we focus on individual fields of scholarship, the concept takes on many, very different meanings. Of primary importance in establishing the concept of 'interactivity' in this case, is how the term is understood in the academic fields of: 1) sociology, 2) communication studies, and 3) informatics [see also Goertz 95] What does sociology's concept of 'interaction' look like? [Duncan 89] writes: »interaction occurs as soon as the actions of two or more individuals are observed to be mutually interdependent«, i.e. »interaction may be said to come into being when each of at least two participants is aware of the presence of the other, and each has reason to believe the other is similarly aware«, in this way establishing a »state of reciprocal awareness«. The basic model that the sociological interaction concept stems from is the relationship between two or more people who, in a given situation, mutually adapt their behavior and actions to each other. The important aspects here are that clear-cut social systems and specific situations are involved, where the partners in the interaction are in close physical proximity, and 'symbolic interaction' is also involved. In other words, a mutual exchange and negotiation regarding meaning takes place between partners who find themselves in the same social context. A situation which communication and media studies would call communication. Therefore, in sociology it is possible to have communication without interaction (e.g. listening to the radio and/or watching TV) but not interaction without communication.

2) As regards the concept of 'interaction' in communication and media studies, there is no such clear-cut answer since there appear to be several different concepts of 'interaction' involved. If we look at the dominant trend within current communication and media studies, what might generally be called the 'cultural studies'

Figure 1: Bordewijk & Kaam's matrix for the four communication patterns

1) If information is produced and owned by a central information provider and this center also controls the distribution of information, we have a communication pattern of the *transmission* type. This is a case of one way communication, where the significant consumer activity is pure reception. Examples would be classic broadcast media such as radio and TV (but also, e.g., live broadcasts of conferences, TV, or multimedia via the MBone). 2) If the exact opposite occurs and information is produced and owned by the information consumers who also control distribution, we have a *conversation* pattern of communication. This is a case of traditional two way communication, where the significant consumer activity is the production of messages and delivery of input in a dialog structure. Typical examples would be the telephone, e-mail, newsgroups, IRC, etc. 3) If information is produced and owned by an information provider, but the consumer retains control over what information is distributed and when, it is a *consultation* communication pattern. In this case, the consumer makes a request to the information providing center for specific information to be delivered. Here the characteristic consumer activity is one of active selection from available possibilities. Typical examples would be various *on-demand* services or on-line information resources such as FTP, Gopher, WWW etc. 4) Finally, if information is produced by the information consumer, but processed and controlled by the information providing center, we have a *registration* communication pattern. In this communication pattern the center collects information from or about the user. In this case, the characteristic aspect is the media system's storage, processing, and use of the data from or about the user. Typical examples would be various types of surveillance, registration, and logging of computer systems.

Among these four information patterns, *transmission* is the only one that is characterized by one way communication. In other words, there is no back-channel that makes an information flow possible *from* the information consumer *to* the media system. Until now, communication and media studies has primarily based its models and insights on the transmission pattern because of the dominant role played by mass communication research. Current media developments including the arrival of 'new media' such as the Internet, intranets, networked multimedia, WWW, etc. have been more or less singularly characterized by a movement away from the transmission pattern toward the other three media patterns. These new media, which open up the possibility for various forms of input and information-flow from information consumers to the system, can hardly be described using traditional one way models and terminology. Seen from this perspective, it might well be claimed that as developments proceed, existing media theory is increasingly less able to explain current media phenomena. Perhaps for these reasons, among others, the established communication research community has developed blind spots in relation to new interactive media. This general problem can only be mentioned briefly here, as we proceed to follow another, more specific trail …

'Interactivity'–The Background Behind the Concept

As [Jäckel 95], among others, has pointed out, the concept 'interactivity' extends–perhaps not surprisingly–from the concept of 'interaction'. A concept which generally means: 'exchange', 'interplay', 'mutual influence'. However, if we focus on individual fields of scholarship, the concept takes on many, very different meanings. Of primary importance in establishing the concept of 'interactivity' in this case, is how the term is understood in the academic fields of: 1) sociology, 2) communication studies, and 3) informatics [see also Gow 96]. 1) What does sociology's concept of 'interaction' look like? [Duncan 89] writes: »interaction occurs as soon as the actions of two or more individuals are observed to be mutually interdependent«, i.e. »interaction may be said to come into being when each of at least two participants is aware of the presence of the other, and each has reason to believe the other is similarly aware«, in this way establishing a »state of reciprocal awareness«. The basic model that the sociological interaction concept stems from is the relationship between two or more people who, in a given situation, mutually adapt their behavior and actions to each other. The important aspects here are that clear-cut social systems and specific situations are involved, where the partners in the interaction are in close physical proximity, and 'symbolic interaction' is also involved. In other words, a mutual exchange and negotiation regarding meaning takes place between partners who find themselves in the same social context. A situation which communication and media studies would call communication. Therefore, in sociology it is possible to have communication without interaction (e.g. listening to the radio and/or watching TV) but not interaction without communication.

2) As regards the concept of 'interaction' in communication and media studies, there is no such clear-cut answer since there appear to be several different concepts of 'interaction' involved. If we look at the dominant trend within current communication and media studies, what might generally be called the 'cultural studies'

tradition, one recurring trait is that the term 'interaction' is used as a broad concept that covers processes that take place between receivers on the one hand and a media message on the other. [Iser 89] actually wrote an essay entitled »Interaction Between the Text and the Reader«. He began by claiming that »Central to the reading of every … work is the interaction between its structure and its recipient«. In brief, his approach is that the work can neither be reduced to the author's text nor the reader's subjectivity, but must be found somewhere between these two poles. And if »the virtual position of the work is between the text and the reader, its actualization is clearly the result of an interaction between the two«. It seems fairly obvious that this is not 'interaction' in the sociological sense. What's missing is genuine reciprocity and an exchange between the two elements involved in that the text can naturally neither adapt nor react to the reader's actions or interpretations. The concept of 'interaction', as it is used here, seems to be a synonym for more noncommittal terms such as 'relationship', 'interpretation' or 'reading' etc. There are, however, also traditions within media and communication studies, where use of the concept of 'interaction' comes closer to the sociological meaning, such as: research in interpersonal communication, research in para-social interaction, traditional media sociology, the 'two-step flow'-model, 'uses and gratification' studies, symbolic interactionism, etc. To review then, it can be noted that the concept of interaction in media and communication studies is often used to refer to the actions of an audience or recipients in relation to media content. This may be the case *even though* no new media technology is being used which would open up the possibility for user input and two way communication; *even though* the social situations are (often) not characterized by the physical presence of an interactive partner; and *even though* the social situations are (often) not characterized by reciprocity and the exchange or negotiation of a common understanding. This is why we cannot speak of interaction in the strictly sociological sense.

3) How is the informatic concept of 'interaction' constructed? The basic model which this concept uses as its starting point is contrary to the sociological tradition the relationship *between people and machines* which in this tradition is often called human-computer interaction (HCI) or man-machine interaction. Historically, this terminology originates from the transition from batch processing, where a large amount of data or programs were collected before being processed by a computer, to the so-called 'dialogue' function, where it was possible for the user to observe partial results, menu choices and dialog boxes and thereby continually influence the performance of the program via new input in–what came to be called–an 'interactive mode'. 'Interaction' in the informatic sense, refers, in other words, to the process that takes place when a human user operates a machine. However, it doesn't cover communication between two people, mediated by a machine,–a process often referred to as *computer mediated communication* (CMC). Within informatics then, (in contrast to sociology) it is possible to have (human-machine) interaction without having communication, but not (computer mediated) communication without also having (human-computer) interaction.

In summary, it can be said that while 'interaction' in the sociological sense refers to a *reciprocal* relationship between two or more people, and in the informatic sense refers to the relationship *between people and machines*, in communication studies it refers, among other things, to the relationship between the text and the reader, but also to reciprocal human actions and communication associated with the use of media as well as (para-social) interaction via the media. Obviously, as far as the concept of interaction is concerned, there is already considerable confusion.

But now let's start to track the concept of 'interactivity'. While sociology doesn't usually use the derivative 'interactivity', the concepts of 'interaction' and 'interactivity' in informatic and media studies appear to be synonymous. In this sense, the concept 'interactivity' or the combination 'interactive media' is most often used to characterize a certain trait of new media which differs from traditional media. The question is, which trait is it?

'Interactivity': Prototype, Criteria or Continuum?

Taking a look at the collection of existing definitions of 'interactivity' spread throughout media studies and computer science, it seems that there are three principle ways of defining the concept: 1) as prototypic examples; 2) as criteria, i.e. given features or characteristics that must be fulfilled, or 3) as a continuum, i.e. as a quality which can be present to a greater or lesser degree.

1) A representative of the first type–definition by prototypic example–can be found in [Durlak 87] »A Typology for Interactive Media«, where among the introduction's qualifying definitions it says: »Interactive media systems include the telephone; 'two-way television'; audio conferencing systems; computers used for communication; electronic mail; videotext; and a variety of technologies that are used to exchange information in the form of still images, line drawings, and data«. This type of definition is, by it's very nature, never very informative, partly because it doesn't point out which traits qualify a given media as interactive or which aspects

connect them. As seen here, and in upcoming examples, the concept of 'interactivity' refers both to media patterns of the consultational and the conversational type. It also becomes clear that the concept of interactivity, understood in this way, is related to the sociological concept of 'interaction' (in the form of the conversational communication pattern) and borrows from the informatic concept of interaction (in the form of the consultation communication pattern).

2) Examples of the second type of definition–interactivity defined as criteria–can be represented, f.ex., by [Carey 89] who suggests the following for the keyword 'interactive media': »Technologies that provide person-to-person communications mediated by a telecommunications channel (e.g., a telephone call) and person-to-machine interactions that simulate an interpersonal exchange (e.g., an electronic banking transaction)«. The last example is explained in more depth a little further on: »most of the content is created by a centralized production group or organization«, and »individual users interact with content created by an organization«. This conceptual construction points more or less directly toward the conversational media type and the consultational media type respectively (and as a result, at the sociological and informatic concepts of interaction) which collectively make up 'interactive media'. Once again there is a certain vagueness to the definition of the concept. More problematic perhaps, is the fact that the definition also excludes services based on the transmission pattern, such as teletext, datacasting, near-video-on-demand etc., which make up the bulk of some TV systems so-called 'interactive services'. Carey himself seems aware of the problem and asks the question whether or not it is possible to draw such narrow boundaries. He writes, »Most scholars would not classify as interactive media those technologies that permit only the selection of content such as a broadcast teletext service with one hundred frames of information, each of which can be selected on demand by a viewer. However, the boundary between selection of content and simulation of an interpersonal communication exchange is not always definable in a specific application or service«. This definition of the concept has the same weaknesses as the majority of other criteria based definitions: the tendency to exclude various media which are generally considered interactive and an inability to use the definition to differentiate between various forms and levels of interactivity.

3) The third possibility, which solves some of these problems is to define interactivity not as criteria, but rather as a continuum, where interactivity can be present in varying degrees. One possible way to structure this type of definition is to base it on the number of dimensions it includes, so that we could speak of 1-dimensional, 2-dimensional, 3-dimensional ... and n-dimensional interactivity concepts.

One relatively simple model of interactivity as a continuum, which operates from only one dimension, can be found in the writing of [Rogers 86]. Rogers defines 'interactivity' as »the capability of new communication systems (usually containing a computer as one component) to 'talk back' to the user, almost like an individual participating in a conversation«. And–a bit farther down–»interactivity is a variable; some communication technologies are relatively low in their degree of interactivity (for example, network television), while others (such as computer bulletin boards) are more highly interactive«. Based on this definition, Rogers creates a scale, in which he lists 'degrees of interactivity' for a number of selected communication technologies on a continuum from 'low' to 'high'. Here, he primarily refers to the concept of 'interactivity' within the consultation pattern. The basic model is clearly 'human-machine interaction', understood in the context of interpersonal communication ('talking back'). It is also because of this consultational aspect (selection available between channels and programs) that classical transmission mass media such as TV and radio can be considered 'interactive'–although to a lesser degree. As is apparent, this attempt to sort and define is relatively rough and lacking in information–a trait that is intensified by Rogers' failure to deliver explicit criteria for the placement of each media.

[Szuprowicz 95], among others, has presented a 2-dimensional concept of interactivity. For Szuprowicz, »interactivity« is »best defined by the type of multimedia information flows«, and he divides these information flows into three main categories. 1) 'User-to-documents' interactivity is defined as »traditional transactions between a user and specific documents« and is characterized by being quite restricted since it limits itself to the user's choice of information and selection of the time of access to the information. 2) 'User-to-computer' interactivity is defined as »more exploratory interactions between a user and various delivery platforms« characterized by more advanced forms of interactivity which give the user a broader range of active choices, including access to tools that can manipulate existing material. 3) Finally, 'user-to-user' interactivity is defined as »collaborative transactions between two or more users« in other words, information flows which make direct communication between two or more users possible. This last form, contrary to the first two mentioned above, is characterized, among other things, by operating in *real time*. Where the first dimension in the matrix is made up of these various information flows, the other is made up of other aspects, which these flows are dependent upon, here again divided into three categories: »access, distribution, and manipulation of multimedia content«. The description indicates that what Szuprowicz calls, 'user-to-user' interaction is related to the sociological concept of interaction, 'user-to-computer'-interaction is related to the informatic concept of interaction, while 'user-to-documents' interaction has an affinity to the interaction concept used by [iser 89]. Along the same lines, the 'user-to-user' information flow is similar to what has been called the conversation communication pattern. The 'user-to-documents' information flow parallels the consultation communication pattern, while the 'user-to-computer' information flow can be said to be a particularly elaborate version of the consultation communication pattern. From this perspective, it also becomes clear that Szuprowicz' differentiation between 'user-to-documents' and 'user-to-computer' is relatively unclear. In most specific cases, it would be difficult to determine whether the 'interactivity' is directed toward a document or toward a platform. The very formulation of the difference appears to refer mostly to the 'degree of manipulability' rather than an actual qualitative difference. This is why the difference is difficult to handle in practice–or to maintain in theory. Instead, this seems to be various forms of the consultation information pattern.

Continuing along the trail to the 3-dimensional concepts of 'interactivity', [Laurel 91] gives us a privileged example. In several contexts, Laurel has argued that »interactivity exists on a continuum that could be characterized by three variables« specifically: 1) »frequency« in other words, »how often you could interact«, 2) »range«, or »how many choices were available« and 3) »significance«, or »how much the choices really affected matters«. Judged by these criteria, a low degree of interactivity can be characterized by the fact that the user seldom can or must act, has only a few choices available that make only slight difference in the overall outcome of things. On the other hand, a high degree of interactivity is characterized by the user having the frequent ability to act, having many choices to choose from, choices that significantly influence the overall outcome – »just like in real life« she adds. As the description of variables indicates, this concept of interactivity moves mostly within the framework of the consultation communication pattern since 'choice' is the recurring term. Understood in this way, the concept can be said to point out three aspects of 'interactivity' within the consultation communication pattern.

An example of a 4-dimensional concept of interactivity, can be found in the writing of [Goertz 95], who simultaneously presents a considerably more elaborate attempt at a definition. After a thorough discussion of various other attempts at definitions, Goertz isolates four dimensions, which are said to be meaningful for 'interactivity': 1) »The degree of choices available«, 2) »The degree of modifiability«, 3) »The quantitative number of the selections and modifications available« and 4) »The degree of linearity or non-linearity«. Each of these four dimensions also makes up its own continuum which Goertz places on a scale. The higher the scale value, the greater the interactivity. Here the 1st dimension falls within what has previously been described as the consultation communication pattern, while the 2nd dimension falls within the conversation pattern. Both the 3rd and the 4th dimensions refer primarily to the possibility of choice and thus fall into the consultation pattern

Finally, there are concepts of interactivity which operate with more than four dimensions, e.g. [Heeters 89] six-dimensional concept of interactivy.

At the End of the Trail?

One possible and reasonably risk-free conclusion from this tracking effort, might well be that the concept of interactivity (as well as the concept of interaction) is outrageously complex and has a long list of very different, specific variations. But it would be unsatisfactory to stop this tracking session with such a disappointing conclusion. In order to arrive at a more satisfactory narrative closure of the quest, a final attempt will therefore be made to suggest a more suitable concept of interactivity, based on the preceding presentations and discussions of the concept.

The above review of the various concepts of interactivity has pointed out, among other things, the inappropriateness of definitions which are based too rigidly on specific historic technologies. It has also pointed out the inappropriateness of defining interactivity via a prototype or as criteria. A definition as a continuum appears to be more appropriate, and at least more flexible, in relation to the many varied levels of interactivity, the many differing technologies and rapid technological developments. It has also become clear that there are different forms of interactivity, which cannot readily be compared or covered by the same formula. There appears to be a particular difference in interactivity which consists of a choice from a selection of available information content; interactivity which consists of producing information via input to a system, and interactivity which consists of the system's ability to adapt and respond to a user. It might, therefore, be appropriate to operate with different–mutually independent–dimensions of the concept of interactivity. And as it may have been apparent from the beginning, or has at least continually been made apparent by this review, the various important aspects of the concept of interactivity can to a great extent be reduced to four dimensions which can be understood using the communication patterns: transmission, consultation, conversation and registration.

Based on this understanding, interactivity may be defined as: *a measure of a media's potential ability to let a user exert an influence on the content and/or form of the mediated communication*. This concept of interactivity can be divided up into four sub-concepts or dimensions which could be called: 1) *Transmissional interactivity*–a measure of a media's potential ability to let the user choose from a continuous stream of information in a one way media system without a return channel and therefore without a possibility for making requests (e.g. datacasting, multicasting, teletext, near-video-on-demand). 2) *Consultational interactivity*–a measure of a media's potential ability to let the user choose, by request, from an existing selection of pre-produced information in a two way media system with a return channel (Gopher, WWW, FTP, *video-on-demand*, on-line information services, etc.) 3) *Conversational interactivity*–a measure of a media's potential ability to let the user produce and input his/her own information in the media system in a two way media system, be it stored or in *real time* (video conferencing systems, news groups, e-mail, maling lists etc.). 4) *Registrational interactivity*–a measure of a media's potential ability to register information from and thereby also adapt and/or respond to a given user's needs and actions, whether they be the user's explicit choice of communication method or the system's built-in ability to automatically 'sense' and adapt (surveillance systems, intelligent agents, intelligent guides or intelligent interfaces, etc.). Since transmissional and consultational interactivity both concern the availability of choice–respectively with and without a request–it is possible to represent them within the same (selection) dimension. The four types of interactivity can then be presented in a 3-dimensional graphic model–an 'interactivity cube'–as attempted in [Fig. 2], which in this form results in 12 different types of interactive media.

I believe that the theoretic approaches presented here are relevant for analyzing and designing networked media and interactive media, and that their relevance for the Internet, intranets etc. will increase in the years to come. Perhaps, more importantly, it is a contribution toward a hopefully greater understanding of the meaning of the concept of 'interactivity' in communication studies and the importance of communication studies to the meaning of the concept of 'interactivity'.

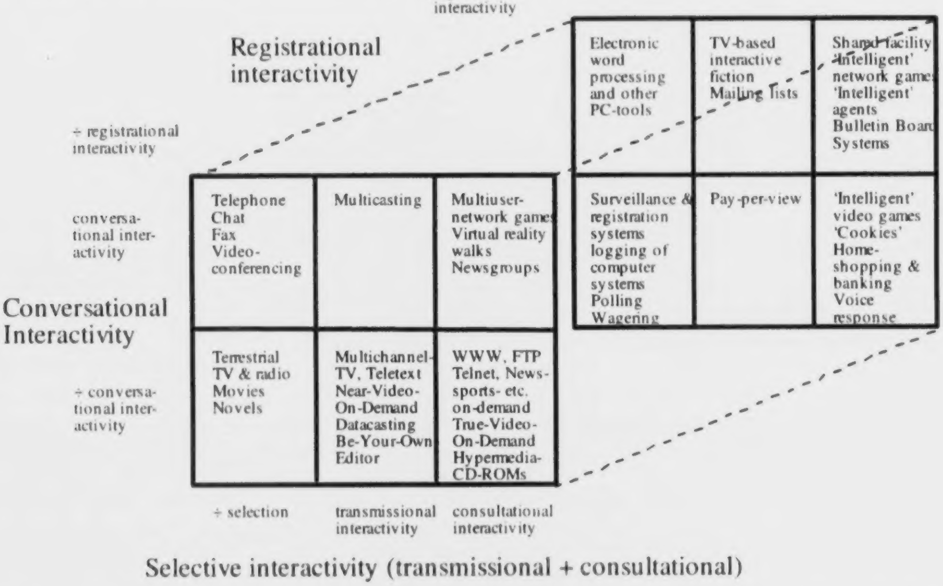

registrational interactivity
Registrational interactivity
÷ registrational interactivity
conversational interactivity
Conversational Interactivity
÷ conversational interactivity
Telephone Chat Fax Video-conferencing
Multicasting
Multiuser-network game Virtual reality walks Newsgroups
Terrestrial TV & radio Movies Novels
Multichannel-TV, Teletext Near-Video-On-Demand Datacasting Be-Your-Own Editor
WWW, FTP Telnet, News-sports- etc. on-demand True-Video-On-Demand Hypermedia-CD-ROMs
Electronic word processing and other PC-tools
TV-based interactive fiction Mailing lists
Shared facility 'Intelligent' network game 'Intelligent' agents Bulletin Boar Systems
Surveillance & registration systems logging of computer systems Polling Wagering
Pay-per-view
'Intelligent' video games 'Cookies' Home-shopping & banking Voice response
÷ selection
transmissional interactivity
consultational interactivity
Selective interactivity (transmissional + consultational)

Figure 2: The 'cube of interactivity': a 3-dimensional representation of the dimensions of interactivity

References

[Bordewijk & Kaam 86] Bordewijk, J., & Kaam, B. (1986). Towards a new classification of teleinformation services. *Inter Media*, 14, 1.

[Carey 89] Carey, J. (1989). Interactive media. *International encyclopedia of communications*. New York: Oxford University Press.

[Duncan 89] Duncan, S. (1988). Interaction, face-to-face. *International encyclopedia of communications*. New York: Oxford University Press.

[Durlak 87] Durlak, J. (1987). A typology for interactive media. In: McLaughlin, M. (ed.). *Communication Yearbook 10*. Newbury Park: Sage publications.

[Gortz 95] Gortz, L. (1995). Wie interaktiv sind Medien? *Rundfunk und Fernsehen*, 4.

[Heeter 89] Heeter, C. (1989). Implications of new interactive technologies for conceptualizing communication. In: Salvaggio, J. & Bryant, J. (eds.) *Media use in the information age: Emerging patterns of a adoption and consumer use*. Hillsdale, New Jersey.

[Iser 89] Iser, W. (1989). Interaction between text and reader. In: Corner, J., & Hawthorn, J. (eds.). *Communication Studies*. London: Edward Arnold.

[Jäckel 95] Jäckel, M. (1995). Interaktion. Soziologische Anmerkungen zu einem Begriff. *Rundfunk und Fernsehen*, 4.

[Jensen 96a] Jensen, J.F. (1996). *Mapping interactive television*. Working papers from the research program: 'The Aesthetics of Television', nr. 15. Aarhus: Aarhus University.

[Jensen 96b] Jensen, J.F. (1996). Mapping the Web. A Media Typology for Information Traffic Patterns on the Internet Highway. *WebNet 96 – World Conference of the Web Society Proceedings. 1996*. Association for the Advancement of Computing in Education, Charlottesville, VA. 219-224.

[Laurel 91] Laurel, B. (1991). *Computers as theatre*. Reading, Mass.: Addison-Wesley.

[Newsweek] Newsweek, May 31, 1993.

[Rafaeli 88] Rafaeli, S. (1988). Interactivity, from new media to communication. In: Hawkins et al. (eds.). *Advancing Communication Science: Merging mass and interpersonal processes*. Newsbury Park.

[Rogers 86] Rogers, E. (1986). *Communication technology. The new media in society*. New York.

[Szuprowicz 95] Szuprowicz, B. (1995). *Multimedia networking*. New York: McGraw-Hill.

Disha: A Direction Toward Software Engineering on the Net.

Raman Kannan
Zhen Chen
William Schroeder
Bo Cao
Rajesh Rao
Joe Occoner

DNA Laboratory
Software Engineering Department
Monmouth University

Veenatejaswini Sanjivi
Kavitha Challapalli
Nanditha Gade

Computer Science Department
Monmouth University

and

Bandish Chitre
Electrical Engineering Department
Monmouth University
West Long Branch, NJ 07764
Email: kannan@moncol.monmouth.edu
http://www.monmouth.edu/monmouth/academic/dna

Abstract: In this article we introduce *Disha*[1] an environment in which software professionals can discover software component from around the world and browse software components in an highly intuitive manner. Disha is comprised of independent softbots that: (1) *perpetually search software sites* on the internet and *catalog* the components found; (2) *proactively offer helpful cues* regarding software components that may be immediately used as is or as a learning aid; and (3) *fetch and present* such software components in an intuitive manner. We use commonly available internet services, such as, simple mail transfer protocol (smtp), file transfer protocol (ftp), hypertext transfer protocol (http), WWW browser and other appropriate internet [Minoli 97] services, indexing software (gais [GAIS URL], glimpse [Glimpse URL]), and interpretive languages such as perl [PERL URL], java [JAVA URL], tcl/tk/expect [TCL URL] to construct Disha environment.

Rationale

The evolution of *internet* as a mainstream communication medium has been super-accelerated with the advent of intuitive interface to the internet (WWW, email and other internet services). Vast amount of useful information assets are made available at ever increasing pace. But, what good is it if only we cannot use, all of it, at will, from anywhere, without having to go through a lot of technical wizardry if not painfully repetetive sequence of conversations with the system (mouse clicks, visiting pages etc...)? In essence: (1) there is a viable infrastructure; (2) vast amount of information assets; but (3) an acute lack of intuitive and effective access mechanisms. The existing access mechanisms do not scale with the size of the content. Our inability to exploit the internet resources has been the motivation behind our efforts in the IDAM project [kannan 96a and 96b]. Our goal, in this project, is to integrate the various tools we have gathered (fabricated here or elsewhere) to:

---

[1] "**Disha**" in Sanskrit means direction. In a sense our objective is to provide direction to willing software engineers. Bhandish and Rajesh coined the term "Disha".

- render the whole internet as a federated reuse repository for software assets;
- facilitate an environment that augments the practioners ability to recall; and
- help practitioners learn by analogies of solutions that may exist somewhere on the net.

While we focus on the internet in this report our ideas, approach and software services can be readily used in IP enabled intranets. Now we present a brief introduction to software construction and maintenance acitivities, operational perspective using use case scenarios (as to how a software professional might use Disha in their construction and maintenance efforts) and an architecture for the Disha Environment.

Software Construction and Maintenance

Software construction (development) and maintenance are two distinct phases within the software development life cycle [See standard text books on Software Engineering for other details]. During the construction phase software engineers implement a design specification in one or more programming languages such that all the details needed to execute the system are available. More often than not software engineers (in sharp contrast to hardward engineers) implement even well known design abstractions from scratch (reinventing the wheel). Consequently the cost to develop software has been increasing while hardware construction costs have been decreasing. Efforts to promote reuse using structured software repositories have not been successful and our premise is that the process of software reuse is perhaps more complicated than necessary. The Disha philosophy is to offer software cues about software components that exist elsewhere. The software engineer need not even be aware of the software component or the location where the software component resides. Thus, with Disha, we posit that software engineers under similar circumstance may now be willing to explore the possibility of using existing assets rather than reimplement them, because:

- (*unobtrusive and intuitive*) the Disha approach does not impose any structure in retrieving or organizing software artificats specifically for reuse.
- (*anticipatory assistance*) Disha offers cues in anticipation, even before the user issues an explicit request; and
- (*transparent navigation and access*) Disha transports such software components from the remote location to the user's environment transparently and without any effort on the part of the user.

Disha in essence lessens the coginitive effort using much less restrictive procedures.

Software maintenance, on the other hand, is the process of evolving a software system after delivery: (1) to meet changing customer requirements; (2) to function in new hardware/software environments; and (3) to fix reported errors. It is estimated that a significant effort is spent on software maintenance compared to any other software life cycle activity and consequently even marginal increase in maintenance efficiency could result in significant reduction in software life cycle costs. However, most software engineering environments are not conducive to software maintenance. Software assets are partitioned based on programming language constructs while the assets are held in flat text files that do not preserve the syntactic and module information of the programming language. So during maintenance human engineers have to translate back and forth from a programming language context to the file layout context. In essence, comprehension, understandability and the ability to recall related software elements with ease are critical capabilities for software maintenance. Disha approach here is to reduce the coginitive effort required to navigate within a software repository exploiting the hypertext transfer protocol (HTTP) and HTML (browser) capabilities of the internet applications.

Now we present the details of the Disha solution from an *operational perspective* -- as to how end users may use the Disha system -- and *architectural perspective* -- as to how the Disha system is being assembled --.

The Disha Solution: Operational Perspective

The idea behind Disha is best explained using two use case scenarios centered around: (1) reuse -- Disha augmented software synthesis -- ; and (2) software maintenance -- Disha augmented software navigation and recall.

Disha Augmented Software Synthesis

Imagine a software engineer composing a solution using a web enabled editor. Let us assume that the design calls for opening a file using the native OS services -- the *open* system call--. As the engineer enters the symbol *open*, Disha proactively recalls other software components that also employ the *open* system call and presents visual cues. Disha visual cues are blinking folder icons that represent software components and these visual cues may be pursued to learn how to use such a function, by a single mouse click, in a user friendly manner. The user if not interested may simply ignore the visual cue offered by Disha and the visual cues fade out in time. However, when the user is not familiar with the -- *open* -- system call, the user may choose to pursue the visual cues offered by Disha. In other words, whenever, the user enters a program construct that is external to the language definition, typically a function name, Disha agents, search an existing database for software resources which include such a function. If a software resource is found a visual cue is offered as a HTML link. If the user pursues the visual cue then the resource is retrieved from the network and converted into HTML format so that the engineer can intuitively browse the software component. The engineer can now understand the exact conditions and the interface protocol with which the desired function is used. Furthermore, if appropriate the engineer may choose to incorporate the discovered resource. We call this --just in time-- reuse.

Disha Augmented Software Maintenance

The ability to browse/navigate in hypermedic space is particularly significant for software artifacts [Pankaj 87] and associated processes because they are intricately connected in non-intuitive ways. The usefulness of semantic information and being able to discover semantic relationship is presented in [Heiler 95]. The need for analyzing the static dependencies amongst software components and the usefulness of such static analysis tools are presented in [Chen 95]. In this context, imagine an engineer, trying to understand some segment of code. For example the engineer may need to examine how a function is implemented or how an instance of abstract data type is composed in the program language specification. Engineers have to contend with the limited support available in some text editors using primitive key board manipulation. In Disha we translate source code into HTML pages such that users can navigate from one location where a variable is used to a location where it is defined and from the point of variable definition to a point where the type declaration is specified. These may be located across files and HTTP/HTML protocol in combination afford an acceptable level of transparency. It should be noted that there are severaal other language specific translation utilities [FILTER URL] on the internet and Disha is one other translation utility. Disha differs in that Disha includes other services seamlessly integrated with the translator utility.

The Architecture

Disha is an instance of the IDAM architecture [kannan 96a] and is comprised of several independent agents each specializing in a specific aspect of the problem at hand as shown in [Fig. 1].

The Disha system builds a catalog of software components in public domain packages. As software engineers work with source files, the Disha matchmaker agents, asynchronously, search for matching catalog entries whenever the engineer enters a symbol external to the language system. In other words ketwords within a language are ignored. If appropriate catalog entries are found, the Disha GoGetters fetch the package from the location indicated in the Disha DB while the Disha MatchMakers present visual cues for each of the catalog entries found on the engineers desktop. If the engineer should pursue a Disha cue, the corresponding packages are first unpacked then translated into html documents. The engineer is then transported to the html document.

GoGetters

GoGetters armed with the meta information on plausible software artifacts of interest fetch them using anonymous ftp and or emailftp. Then the GoGetters disect the downloaded packages and pass them as a collection of source files to *FormatManagers*.

Discovery Agents

These agents visit various sites and examine available artifacts, extract meta-information and catalog them. To begin with location, package information, language, abstractions and meta information such as timestamp and size are cataloged.

Summary

In this article we have presented the rationale for just in time reuse and maintenance using web technologies. We have presented the rationale and the architecture for Disha, an environment which encourages just-in-time reuse and static analysis of software components using WWW technologies. In essence we believe that hypermedia and internet technologies, in particular, will radically change the way we practice software having an impact on every aspect of software life cycle. Disha is a small step in that direction.

Reference

[Minoli 97] Daniel Minoli. (1997). *Internet Engineering: Technologies, Protocols and Applications.* McGraw Hill.

[GAIS URL] Gais Team, Internet ReSearch Laboratory, Department of CSIE, National Chung Cheng University, Taiwan. *Grep and Indexed Search. http://gais.cs.ccu.edu.tw/IndexQuery.html*

[Glimpse 94] Glimpse Working Group, University of Arizona. *Glimpse: A tool to search entire file systems. http://glimpse.cs.arizona.edu:1994/*

[PERL URL] *Practical Extraction and Report Language.* The Perl Language Home Page. *http://www.perl.com/perl*

[JAVA URL] *The Java Language: An Overview. http://www.javasoft.com/doc/Overviews/java/java-overview-1.html.*

[TCL URL] *Tool Command Language/tcl. http://www.sunlabs.com/research/tcl.*

[Kannan 96a] Raman Kannan, (1996). *Semantic Web for Collaborative Software Processes. http://starbase.monmouth.edu/monmouth/academic/dna/weti96ff.htm.*

[Kannan 96b] Raman Kannan, (1996). *I-DAM: An Active Digital Archive Management. http://starbase.monmouth.edu/monmouth/academic/dna/w3cf.html*

[Pankaj 87] Pankaj, K.G., and Scacchi, W., (1987). On designing hypertext systems for information management in software engineering. *Hypertext 87.*

[Heiler 95] Sandra Heiler (1995). Semantic Interoperability:Research Directions in Software Engineering. *ACM Computing Surveys*, June 1995, 271-273.

[Chen 95] Yih-Farn Chen, "Reverse Engineering" in Practical Reusable Software. Editors Balachander Krishnamurthy, Wiley and Sons.

[FILTER URL] A complete list of translators could be found either at
http:// www.w3.or g/pub/WWW/Tools/Filters.html or
http:// www.w3.org /pub/WWW/Tools/Prog_lang_filters.html

Acknowledgements

Disha and other DNA efforts would not have been possible but for the support and encouragement of Mr. William Sweeney, Mr. Byron Nicas, Dr. Vadim Dosychev, Dr. Dan Marcellus, Dr. John Sabini, Mr. Peter DeLucia, Professor Jorge Diaz, Dr Jerry Powell, Mr. Robert Bekefi, Dr. Larry Dworkin, Director Center for Technology Development and Transfer, Professor Harris Drucker, faculty members of the Computer Science Department, Grant in Aid Subcommittee for supporting other WWW related projects, and Concurrent Engineering Research Center, West Virginia University. Most importantly without the graduate student contribution Disha would be nothing more than a vision. *Disha and DNA efforts are not funded by public or private* .

On Developing a Web-Based Operating Environment

Lee Kennedy, Jim Carter
Department of Computer Science
University of Saskatchewan
Canada
email: {lek130, carter}@cs.usask.ca

Abstract: The Web has achieved tremendous popularity in its short history, and has undergone radical changes. The continued addition of features to the Web has transformed it from an information retrieval system to an operating environment capable of running full-scale applications. Users will need help managing this environment if the Web is to keep the ease of use which has made it popular. It is desirable to provide users with a single interface which integrates both local and web-based applications. This paper considers the move towards such an environment, the difficulties of maintaining simplicity, and user requirements of such a system.

1 Introduction

The rapid growth of the Web has perhaps only been matched by its rapid changes in functionality. From its inception as a hypertext document retrieval system the Web has grown to incorporate fill-out forms, database gateways, CGI scripts, and even animation. The use of Java, Javascript and third-party plug-ins now provides the potential for an almost unlimited array of new features. However, these new features have done much more than simply improve the variety of documents available for retrieval over the Web. These additions have fundamentally altered the way in which the Web is used.

[Fig. 1] shows the progression of change in the Web, and the corresponding use paradigm. The Web was developed as a system for accessing hypertext documents. This led to the large-scale development of Web-specific documents. At this stage the Web was used as an information retrieval tool. New features were soon added, including: interactive forms, database gateways, and scripting. With these additions web documents moved from passive information repositories to active information consumers. Developers began to create Web applications, rather than simple documents, and the Web became a tool for information processing. Plug-ins, further user interface (UI) improvements, and increased communication between the web browser and other applications (such as the ability to handle email and newsgroups) have further expanded the Web's potential. The web browser, rather than the machine's operating system, becomes the environment in which these web applications run. As web applications become more prevalent and more complicated users will need a method to manage this new environment (similar to the management tools for local applications provided by the local operating system). Further improvements in the Web provide the opportunity for the next step forward with the integration of local and web applications. This would transform the web from an information processing tool into a full user operating environment.

Each of its developments has made the web more complex in terms of its functionality, while retaining its interfacing simplicity and ease of use. As new functionality is continually added there is a real danger of losing this simplicity and destroying a crucial factor to the success of the web. There is little doubt that new functionality will continue to be added, but there are important HCI challenges which must be addressed if users are to accept these changes. Thus the important issue in moving to the web as operating environment stage is not "Can the web be made to encompass both local and web applications?", but rather "*How* can the web encompass both local and web applications *while still maintaining it's simplicity from a user's perspective?*"

2 Web Applications Are Being Developed

One of the primary assumptions involved in integrating local and web applications is that web applications will actually exist and users will want to use them. Studies show that application research is ongoing in two major areas::

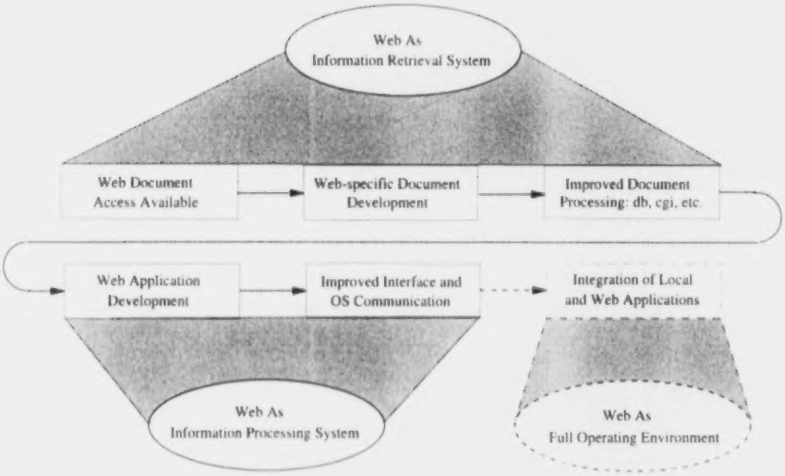

Web As
Information Retrieval System
Web Document
Access Available
Web-specific Document
Development
Improved Document
Processing: db, cgi, etc.
Web Application
Development
Improved Interface and
OS Communication
Integration of Local
and Web Applications
Web As
Information Processing System
Web As
Full Operating Environment

Figure 1: Progression of Change in the Web

- integrating existing legacy systems with a web interface [Giradot et al., 1995] [Perrochon, 1995]
- developing new systems using web technology [Cutkosky et al., 1996] [Katz and Silva, 1994]

These studies all relate to large-scale business systems. This research shows that at least some users will encounter web applications in the workplace.

There are also numerous applications aimed at recreational users. Many sites feature searching and filtering applications, and provide "shopping carts". Sports sites like ESPN's Sportszone and CBS's Sportsline provide real-time scoreboards which scroll sports scores and are customizable by the user.

The combination of large-scale business application development and the rising proliferation of smaller, "recreational" applications makes it increasingly likely that users will find web applications which they either must or want to use.

3 Challenges to Retaining a Simple Interface

With web applications on their way, users need a way to manage these applications while still maintaining the simplicity of the web interface. Research in user interface design and distributed systems identify some of the challenges to retaining interface simplicity.

3.1 User Interface (UI) Challenges

It is imperative that evolving to a web-based operating environment avoid incorporating the known pitfalls of bad interfaces. Barfield(1993) describes three main types of bad interfaces:

- *minimal UI* - user has unsupported tasks, no options, etc.
- *complicated UI* - user is overwhelmed by "featuritis"
- *good-looking UI* - user is enticed by flashy UI with little actual functionality

If the web interface does not adapt to the changing nature of the web it will move into the "minimal UI" category. If too many new features are added the interface will fall into the "complicated UI" category. Additions and modifications of the interface must be made in a way consistent with the current interface.

Card(1989) presents four pressures which make it increasingly difficult to provide users with a good UI:

- increased functionality of systems
- more cognitive tasks being performed
- applications becoming more complex
- UIs are expected to be more interactive

BEST COPY AVAILABLE

Each of these pressures is at work on the web interface, and will become stronger as the use of web applications increases. These pressures make it harder to avoid the extremes of Barfield's "minimal UI" and "complicated UI".

Card's pressures show that it is increasingly difficult for users to understand communication between the system and the user. Semiotics deals with how shared meaning is conveyed, and is related directly to UI design by Mullet and Sano(1995). An interface contains uses various signs to communicate meaning to the user. For effective communication to take place signs must work on three levels:

- *syntactic* - the relationship between parts of the sign
- *semantic* - the relationship between the sign and the actual object
- *pragmatic* - the way in which the sign is actually interpreted

As systems become more complicated it is difficult to select UI elements which have a clear semantic relationship to the task they represent. As the semantics of an icon becomes less clear there is a greater chance that the user will misinterpret the sign, a failure on the pragmatic level. A web-based service environment must not violate the shared meaning already created with the user without careful thought.

3.2 Distributed Systems Challenges

The Web is a distributed system of information and applications. This distributed nature is responsible for many problems involved with data navigation and visualization. Traditional distributed systems research has dealt with some of these issues and can provide some guidance.

One of the issues is how to provide users with applications on a large number of machines without forcing them to deal with the underlying technical details. There are three main elements to providing an application to the user:

- application data storage
- application code storage
- resources for execution (memory, CPU cycles, etc.)

In a normal local application all of these elements are provided locally. The introduction of a distributed system complicates these elements by providing three different locations for each:

- local
- remote
- hybrid - partially local and partially remote

There can be different levels of remoteness. A remote machine may be part of a local intranet or may just be part of the internet. Hybrid and remote locations may consist of a single machine or multiple machines.

The Legion system [Grimshaw et al., 1997] provides a simple solution. Legion users deal with a single "virtual machine". Users are unaware of whether an application is local or distributed over remote machines. Legion provides complex functionality without the user *needing to be aware of the complexity*. A similar type of complexity hiding will be required by a web-based operating environment.

There is a trade-off however between simplicity and informed decision-making. At times users will need to be aware of details to make informed decisions. Particularly users need to:

- *form response time expectations* - applications with remote elements are sensitive to network traffic loads, bandwidth restrictions, etc.
- *estimate local resource consumption* - local disk storage, CPU load, etc.
- *form stability estimates* - internet applications have no guarantees of availability or maintenance; intranet applications will likely have better support; local applications are controlled by the user
- *estimate application trust* - users must be aware that web applications include unreliable data or potentially dangerous programs. This is even more important for the web, where any machine may be involved, than with distributed systems where only known machines are involved.

4 Developing an Environment Which Meets User Needs

The primary goal of any system should be to meet the needs of its users. A new system must go above and beyond the existing system by improving the service of previously supported needs, supporting new user needs, or both.

In developing requirements requirements for a web-based operating environment there are a number of user-oriented questions which must be answered:

- Why would users want this new system?
- What do users need to do with the system?
- What will users need to help them do these things?

The answers to these questions form the basis of a requirements analysis for the new system.

4.1 Why would users want this new system?

Users must see the potential for them to benefit if they are to undertake the effort of learning a new system. If a system does not provide the expected benefits to the user, it will fail. The key user benefits of a web-based operating environment will be to:

- *reduce and simplify the user interaction with the local OS interface:* Many current OS interfaces are complicated and obscure. New users are often intimidated by this complexity. With the growing popularity of the Web, many users know how to use their web-browsers, but can perform few other computer tasks. A new operating environment must provide users with an easier way of handling local applications.
- *provide management functionality for web-based applications:* Currently users have no method of managing web-based applications. Web browsers rely on the concept of static web documents and use tools such as bookmark files and forward/back buttons. These tools do not provide the functionality for handling full web applications. Users need methods for handling these new web applications.
- *manage local and remote applications in the same manner:* One of the crucial requirements of a web-based operating environment will be to integrate the management of both types of applications.
- *provide user with the same environment across different hardware and OS platforms:* Users are faced with different interfaces and system behaviour as they switch hardware and software platforms. Meeting this requirement allows users to perform tasks using the same interface on a variety of platforms.

A new operating environment is required to continue to provide the following services to the user:

- *management of local applications:* Users may not like the OS interface for managing local applications, but the basic functionality must still be retained in some form.
- *current web-browser services:* No major changes should be made in how users access these services. The simplicity and usability of the Web interface is popular. Changes may alienate a large segment of users.

4.2 What do users need to do with the system?

The required user tasks can be divided into four basic groups by the type of entity being dealt with:

- *managing the application space:* The set of applications available to the user form the application space. The specific applications within the space and their relationship to one another define the shape of this space. The space is managed by adding and removing applications from the system. Users need to mentally visualize this space by defining relationships between applications, and must be able to group applications, and move them from one group to another. Users will need to manage different aspects of the space at different times. The three main levels of the space are:
    - *physical* - the physical location of applications in system
    - *logical* - the logical groupings of applications which the user has defined
    - *visual* - the visual organization the user has chosen for the logical layout

The interaction between these levels is complex. Two applications may be logically grouped and yet have no physical correspondence in terms of where they are stored. For an application to be included at the logical level it must be included in the physical level. Users must be able to manage all three of these levels: the visual to ensure an intuitive and appealing interface; the logical to provide an environment consistent with the user's mental understanding of the system; the physical to control physical resources such as disk drives.

- *managing individual elements:* Users must manage individual elements in the application space. They must be able to select visual representations(icons) for applications and groups of applications. The application space does not change: the set of applications and their relationships to one another are the same. Other tasks include setting application control flags and naming applications or groups.
- *managing the environment space:* Some tasks affect the overall user environment without altering the application space, such as color, font type and size, and default settings. These types of tasks help the user create an environment tailored to their needs.
- *managing the execution of applications:* Users must be able to start, stop, and switch among running applications, as well as monitor how the execution is progressing, and the resources being consumed. A running application will be called an *invocation* to distinguish it from the application code.

4.3 What do users need to help them use the system?

A good system must not only provide basic functionality, but also help the user take advantage of the available functionality. Users take advantage of functionality through a three-phase interaction cycle:

- *observe* the current state of the system
- *understand* what is happening in the system based on these observations
- *control* the system by acting on this understanding

The key phase in helping the user is understanding. Understanding allows users to make informed decisions. Observation and control must be provided in a way which facilitates understanding. For users to have true understanding they must have:

- *comprehension of observations* - what is going on?
- *awareness of available control options and their results* - what can I do?
- *knowledge of how to exercise control* - how do I do what I want?

There are a number of requirements which help the user understand their environment:

- *appropriate level of detail:* There is a tradeoff between simplicity and control. Some users have simple application use patterns. They do not wish to be overwhelmed with detailed observation data about the environment or a large range of control actions. Other users want fine-tuned control and with detailed data and more complicated control tasks. Different users require different sets of observation and action tasks from the environment.
- *appropriate user guidance:* Timely feedback with appropriate detail, warnings of potentially dangerous decisions, and advice on interpreting observations and selecting actions help the user understand the environment. Users should be able to select among sets of defaults to personalize the user guidance.
- *help system:* The help system must include context-sensitive help and general system help. Several levels of help will be necessary to provide details appropriate to the knowledge level of the user. Context-sensitive help must be automatic, and all help must be available on-demand.
- *ease of use:* A simple, intuitive, easy-to-use interface is the best way to help the user. If users cannot quickly learn at least a minimal set of functions to be productive they are unlikely to adopt the system.

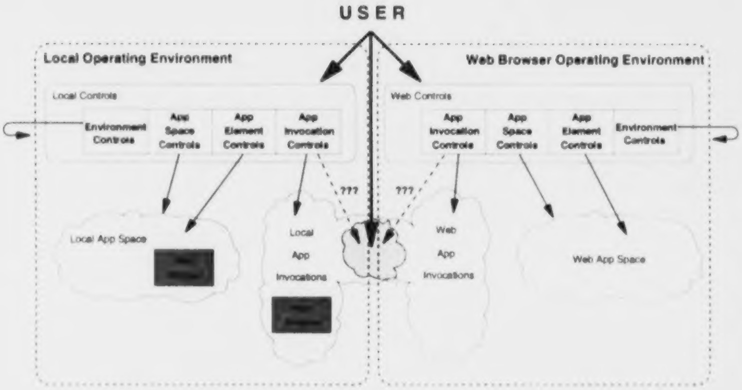

USER
Local Operating Environment
Web Browser Operating Environment
Local Controls
Environment Controls
App Space Controls
App Element Controls
App Invocation Controls
Web Controls
App Invocation Controls
App Space Controls
App Element Controls
Environment Controls
???
???
Local App Space
Local App Invocations
Web App Invocations
Web App Space

Figure 2: Control Flow in Current User Operating Environment

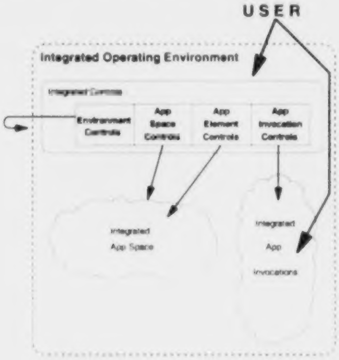

USER
Integrated Operating Environment
Integrated Controls
Environment Controls
App Space Controls
App Element Controls
App Invocation Controls
Integrated App Space
Integrated App Invocations

Figure 3: Control Flow in an Integrated User Operating Environment

5 Integrating the User's Environment

Integrating local and web-based applications into a single operating environment will alter the way users interact with their system. [Figure 2] shows the current user interactions where local and web-based applications are separate. The environment consists of two disjoint application spaces, and two different sets of controls. The local environment is controlled by the operating system controls while the web environment is controlled through the web-browser. The browser itself is a local application, and runs as a local invocation. Web-browser controls are currently limited by the "document" metaphor used by most browsers (web "pages", bookmarks, etc). Some web applications, such as Java applications, fall into the overlapping area between local and web-based invocations. Such applications involve both remote and local elements and it is not clear whether they are handled by local controls, remote controls, or both. This is a potential source of confusion for users.

[Figure 3] shows an integrated environment where a single set of controls is used for both local and web-based applications. This integration allows the user to work with a single application space and a single set of controls for all aspects of the operating environment. In both operating environments, the user is shown interacting directly with application invocations, representing the use of the application interface.

An integrated environment will need to provide some of the functionality of the current local and web-browser control sets. There are three main ways this can be accomplished:

- as a new stand-alone application interface with OS and web browser functionality
- as the OS interface with additional web browser functionality
- as the web browser interface with additional OS functionality

Using a new application requires the user to learn another interface and duplicates both OS and web browser functionality. Many users are already intimidated by the complexity of current OS interfaces,

BEST COPY AVAILABLE

and adding additional functionality will only make the problem worse. The OS interface also eliminates the potential for platform independence, one of the potential benefits of an integrated environment.

Of the three alternatives, using a web browser interface offers the most promise for meeting user needs. Browsers such as Netscape already provide a known, easily-learned interface across multiple platforms.

6 Conclusion

A web-based operating environment does not require any radical new ideas or revolutionary technologies. Recognizing the impact of the developments which have occurred and creating a synthesis of the right ideas reveals a new and exciting possibility: a platform-independent operating environment providing users with control over both local and web-based applications, with a familiar and widely accepted interface. The success of such an environment will hinge on whether the evolution of the web can continue while still maintaining the HCI characteristics which have made it popular with users. The progress of the web has gone far beyond what was anticipated by its creators. As the web diverges more radically from its origins it becomes increasingly difficult to incorporate these changes gracefully for the user. HTML has been fundamental to the success of the web, but the movement from web documents to web applications exceeds its capabilities. Dealing with the above requirements requires the browser to deal with applications as objects. The browser must be able to interact with and manipulate these objects. There is a need for an applications mark-up language(AML) to augment HTML. Any formulation of an AML must take into account these requirements.

References

[Barfield, 1993] Barfield, L. (1993). *The User Interface: Concepts and Design*. Addison-Wesley.

[Card, 1989] Card, S. K. (1989). Human factors and artificial intelligence. In Hancock, P. and Chignell, M., editors, *Intelligent User Interfaces: Theory, Research and Design*. Elsevier Science Publishers.

[Cutkosky et al., 1996] Cutkosky, M. J., Tenebaum, J. M., and Glicksman, J. (1996). Madefast: Collaborative engineering over the internet. *Communications of the ACM*, 39(9):78–87.

[Giradot et al., 1995] Giradot, J., Michel, B., and Phillippe, J. (1995). W3 as the ultimate integration tool. http://zenon.inria.fr/ERCIM/W4G/positions/Giradot.html. ERCIM/W4G International Workshop on WWW Authoring and Integration Issues, Paris.

[Grimshaw et al., 1997] Grimshaw, A. S., Wulf, W. A., and the Legion Team (1997). The legion vision of a worldwide virtual computer. *Communications of the ACM*, 40(1):39.

[Katz and Silva, 1994] Katz, R. H. and Silva, M. J. (1994). The case for design using the world wide web. Technical report, UCBerkeley. ftp://tr-ftp.cs.berkeley.edu/pub/tech-reports/csd-94-837/all.ps.

[Mullet and Sano, 1995] Mullet, K. and Sano, D. (1995). *Designing Visual Interfaces: Communication Oriented Techniques*. Sunsoft Press.

[Perrochon, 1995] Perrochon, L. (1995). W3 middleware: Notions and concepts for web access to legacy data. ftp://ftp.inf.ethz.ch/doc/papers/is/ea/4www95.html.

Content is Key:
Viewing Conceptualized Content

Robert E. Kent
School of EECS
Washington State University
Pullman, WA, USA

rekent@eecs.wsu.edu

Christian Neuss
Fachbereich Informatik
Technische Hochschule Darmstadt
Darmstadt, Germany

neuss@informatik.th-darmstadt.de

Abstract: Conceptualized content is key. In this paper we provide a new conceptual approach for accessing and organizing information. This approach is based upon a mathematical theory of conceptual knowledge and a graphical metaphor for interactive exploration over the conceptual content of networked information.

Searching the Web

With the Web well into its 7th year, resource discovery still is the biggest and most serious problem for users. Search engines such as Webcrawler, Alta Vista, Excite, InfoSeek, Lycos, and Yahoo are currently very popular, but can be frustrating to use. As the Internet grows, the sheer vastness of the information space will make free text search engines less and less useful. In addition, quite a bit of material on the Web is inherently inaccessible to robots/crawlers, either due to robot exclusion, or because the data is accessed through a dynamic mechanism such as a database search form.

Users frequently complain about poor service on the Web. For the users experience to improve markedly, several problems need to be addressed. These include diverse information quality, little semantics, low precision in search results, and little support for domain specialization.

We will now take a closer look at the drawbacks involved with robot technology based free text searching.

The Web: A Digital Wasteland?

Many critical voices characterize the World Wide Web as a "vast intellectual wasteland." Indeed, while there is a lot of high quality information on the Web, identifying and finding it is still a difficult and time-consuming task. This is in part due to the fact that the Web was designed without a built-in search facility. But other factors contribute as well.

Consider, for example, a public library: the material is well organized into a hierarchy of categories, and can be searched by using a variety of meta-information items such as keywords, publication date, or author. People who know how to "organize" information - librarians and other information specialists, assign this information. Since all these documents are reviewed, a certain minimum level of quality can be expected.

Unlike the public library, the Web lacks categorization of content, and has no filtering mechanism that eliminates information that is irrelevant, outdated, or simply plain wrong. In order to conduct a more meaningful and relevant search, you need a more refined method for conceptually and semantically organizing the gigabytes of digital information.

Spider Based Search Engines

Robotic search engines such as Webcrawler or the popular Alta Vista provide a fast and simple means of performing free text searches on the World Wide Web. By automatically retrieving remote pages, indexing their content, and recursively following the links stored in them, they maintain a database of a large number of Web pages. This database provides Web users with a search facility, often with a very sophisticated query syntax.

Although the numbers claimed by robot maintainers should be taken with a grain of salt, it is safe to say that the number of pages available to robotic indexes is currently on the order of tens of millions of pages. The most tempting aspect of these full-text search engines is the notion that they seem to eliminate the need to organize information in order to be able to find it later. Just enter a number of keywords, push the submit button, and a few seconds later you will be prompted with a number of references to material relevant to your query.

Unfortunately, this doesn't work. Besides the problem of physically accessing a continuously growing amount of information (Already, it takes Alta Vista's crawlers over six weeks for a complete walk of the Web!), there is also the inherent weakness that all free text based methods share. This is the gap between the syntactic content and the actual semantic meaning. When performing a free text search, many documents will be retrieved that contain a desired word, but which may not actually use the word with the desired meaning or in the desired context.

Enriching Contents

The key for overcoming these problems is to create quality Web content. There are two approaches for doing this. One approach uses intelligent information retrieval in order to build huge intelligent programs on the client. This will leave the Web infrastructure as it is. It has less political and non-technical hurdles, such as adopting standards and actually getting people to use these things. The goal of the other approach is to distribute more knowledge into the Web infrastructure. This makes it easier to intelligently browse/query/navigate the information. It is also in accord with the peer-to-peer paradigm, which advocates a movement away from omniscient servers - those that contain all the knowledge.

Although both approaches are viable, and in the long run we understand the need for using both, in the near term we view the second approach to be more effective. In its use of ontologies, it is spiritually close to conceptual knowledge processing (discussed below). This approach has one option that masquerades as a solution but is not, and two genuine and semantically meaningful options.

Keyword stuffing:

Really a non-solution, but widely used. Keyword stuffing is the process of "hit-enhancing" a page by the manipulation of Web page content. Current search engine technology allows Web document authors to manipulate the contents of their Web pages is such a way that the page will be ranked higher in the search engines query results, thus giving the page a higher visibility. This technique is often misused; for example, to draw attention to commercial pages, irrelevant, but popular, keywords are added. In fact, an attempt to give a page a higher rank order can be as simple as adding multiple instances of a keyword (you can actually find pages with thousands of occurrences of "Intranet" all in a row). This goes as far as reverse engineering the ranking algorithm and trying to make use of the sorting algorithm employed.

Needless to say, keyword stuffing is neither a semantically meaningful nor a truthful option. However, the fact that people are willing to go to such great lengths, in order to appear high up in search engine results, gives reason to hope that other semantically more meaningful markup will also be adopted, once the tools for retrieving and browsing this information are in place.

Meta tags:

The HTML meta tag is an invisible HTML element that was designed for augmenting an HTML document with descriptive meta-information. It allows for the specification of an arbitrary number of name/value pairs in the document header. Currently, almost all search engines support META tag markup. For instance, Alta Vista allows keyword definition through the following specification

<meta name = "keywords" content = "Information Retrieval, Resource Discovery, Text Processing">

Ontologies:

What is an ontology? This question can have several different answers: an explicit specification of a conceptualization (Tom Gruber); a description of the concepts and relationships that can exist for a community of agents; ISA hierarchy (taxonomy) plus relationships (SHOE); object typed metadata (Synopsis File System); a concept lattice (WAVE), which is an implicit specification of a conceptualization or better, the conceptualization itself! How are ontologies related to conceptual knowledge processing? Ontologies are specifications of conceptualizations. Concept spaces are "true" conceptualizations. Thus, ontologies specify conceptual space. This is our choice. We are currently using ontological annotations of HTML documents in a SHOE-like format, for the conceptual specification of a community's web of networked information.

Conceptual Knowledge Processing

A formal treatment of information must include some formal understanding of concepts and conceptual relations. Formal concept analysis yields such understanding by mathematizing the philosophical view of a concept as a "unit of thought" having as constituents its extension and its intention. A *formal concept* consists of a collection of objects exhibiting one or more common attributes. The *extent* of a concept is the aggregate of objects that it includes or denotes. The *intent* of a concept is the sum of its unique attributes, which, taken together, imply the concept. Formal concept analysis is based on a order-theoretic model for (formal) contexts from which concepts and conceptual hierarchies can be formally derived. A basic result is that the formal concepts of a formal context always form the mathematical structure of a lattice with respect to the sub-concept /super-concept relation. This complete lattice, called a *concept lattice*, conceptualizes our information space. For ease of reference into this conceptualization, we provide a naming facility via conceptual views. A *conceptual view* names or "bookmarks" a formal concept within a concept lattice. A *concept space* collects together all conceptual views, along with all objects and attributes, and hence is a named part of a concept lattice, or a referenceable conceptualization.

An adequate theory of knowledge is more than just a theory of knowledge representation. According to Rudolf Wille, the founder of formal concept analysis, knowledge is elaborated through inferencing, and knowledge is created and augmented through acquisition. In addition, an adequate theory should provide an approach for the development of knowledge communication tools. The notion of a concept space helps to provide for such a theory. A specification of conceptual knowledge is based upon the 3 fundamental notions of objects, attributes, and conceptual views, which are connected together by 4 basic relationships:

Incidence: An object has an attribute.

Extent: An object is an instance of a conceptual view; conversely and equivalently, the conceptual view has that object in its extent.

Intent: An attribute abstracts from and distinguishes a conceptual view; conversely and equivalently, the conceptual view has that attribute in its intent.

Sub-view: A conceptual view is a subtype of another (super-ordinate) conceptual view

The 3 basic notions and 4 basic relationships of conceptual knowledge form the components of a concept space. Figure 1 represents a conceptual view (the central node) within a concept space with its extent/intent local neighborhood spaces. There are 3 concept spaces here, each represented by a spindle shape: the global space and the extent/intent local neighborhood spaces for the conceptual view.

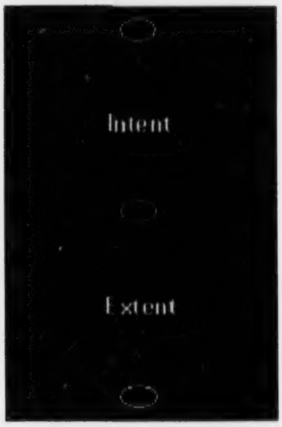


Figure 1: A Conceptual View in a Concept Space

A concept space names part of a conceptual knowledge universe represented by a concept lattice. A conceptual universe forms a context or background concept space within which various user-customizable concept spaces can be created, explored, developed, extended, related, etc. The representational mechanism of a concept space serves as a firm foundation for the basic paradigms of Internet resource discovery and wide-area information management systems: organization-navigation and search-retrieval. The use of concept spaces is a natural outgrowth of the original approach of Formal Concept Analysis for structuring and organizing the networked information resources in the World Wide Web.

Creating a WAVE

The project Creating a WAVE is a multi-year project at Washington State University, which is funded by Intel Corporation. The general goal of the project is the conceptual organization of a community's information space on the World Wide Web. The project will develop an advanced (Networked Information Discovery and Retrieval) NIDR system called WAVE, which fuses the current NIDR system technology with a mechanism for "dynamic distributed classification." Since the Intranet for a commercial company or a university is such a web community, the WAVE system applies directly to the conceptual organization of Intranets.

The project seeks to address the following research question: What is the appropriate architecture for a digital library?" The research goal of the project is to demonstrate in the distributed context of the World Wide Web that the WAVE system, using both the technique of automatic classification and the notion of conceptual space, provides the kernel architecture for a digital library.

The use of ontologies in various knowledge-sharing projects has much in common with the WAVE approach for the conceptualization and sharing of knowledge. An ontological extension to the World Wide Web specifies a conceptualization by the WAVE system of a Web community's information space.

Conceptual Knowledge Markup Language

CKML is an XML application being designed by the WAVE team for use in knowledge and structured metadata representation. CKML seeks to extend existing Web metadata standards with conceptual knowledge information. The ideas developed in CKML come from at least two wellsprings: the SHOE initiative at the University of Maryland and the CKP principled approach to knowledge representation and data analysis being

BEST COPY AVAILABLE

developed by Rudolf Wille's group at the Technische Hochschule Darmstadt.

The **WAVE** project is using CKML to specify a conceptual interface for a variety of information resources, including entertainment information space (movies, television, etc.), corporate information space (Intel press releases), higher education information space (Washington State University), professional society information space (ASIS), and others. We intend to design translation mechanisms into CKML from other knowledge and metadata representation schemes. Currently we have prototype translators SHOE-to-CKML and CDF-to-CKML, we are working on a translation scheme for DC-to-CKML, and we plan to have a translator MCF-to-CKML. More information is available at the Web pages listed in Table 1.

We believe that CKML will influence the development of MCF and other XML metadata proposals, and hence influence the development of XML itself. Once translated into CKML, and thus annotated with conceptual knowledge information, we analyze/visualize the data using a conceptual scaling methodology and the **WAVE** conceptual browser.

CKML = Conceptual Knowledge Markup Language
WAVE = Web Analysis and Visualization Environment
CKP = Conceptual Knowledge Processing
OML = Ontology Markup Language
SHOE = Simple HTML Ontology Extensions
CDF = Channel Definition Format
DC = Dublin Core
MCF = Meta Content Framework
ASIS = American Society for Information Science

Table 1: Links Related to Conceptual Knowledge Markup Language

The WAVE Conceptual Browser

The basic conceptual browsing style is dual mode (both extensional and intentional) but browses only over the global scope, although it displays the local scope. The basic style is illustrated in Figure 2, which corresponds to extensional mode display (the mode of the global scope). This window is partitioned into three panes: focus, display, and definition. The focus ("global scope - extensional mode") pane on the left corresponds to the left window tree hierarchy in Windows File Explorer, and the display ("local scope - intentional mode") pane at the bottom right corresponds to the right window report display in Explorer. The former distinguishes ascending views and intentional attributes (stuff above), and the latter contains only descending views and extensional objects (stuff below). The definition pane at the top right is used to move around the concept lattice by taking suitable meets and joins of contained elements. There is also, of course, the intentional mode display consisting of a "global scope - intentional mode" focus pane containing descending views and extensional objects with other views and objects in the intentional similarity part, and a "local scope - extensional mode" display pane containing ascending views and intentional attributes. The current version of the **WAVE** conceptual browser is downloadable from the Web page http://wave.eecs.wsu.edu/WAVE/versions/versions_2_x.html.

Summary and Future Work

The **WAVE** project has two principal development phases: off-line and on-line. The project has been under way for one and a half years at present. During 1996 the first phase developed an off-line conceptual navigator in order to study various issues of functionality, usability and scalability. During 1997 the second phase is developing an on-line conceptual navigator with a CKML back-end and an ActiveX/Java front-end. In addition to movies, we are using an Intel press release data set as an illustrative demonstration. The latter application illustrates how formal concepts can be used to represent user interest profiles, which are useful for filtering in various push technologies.

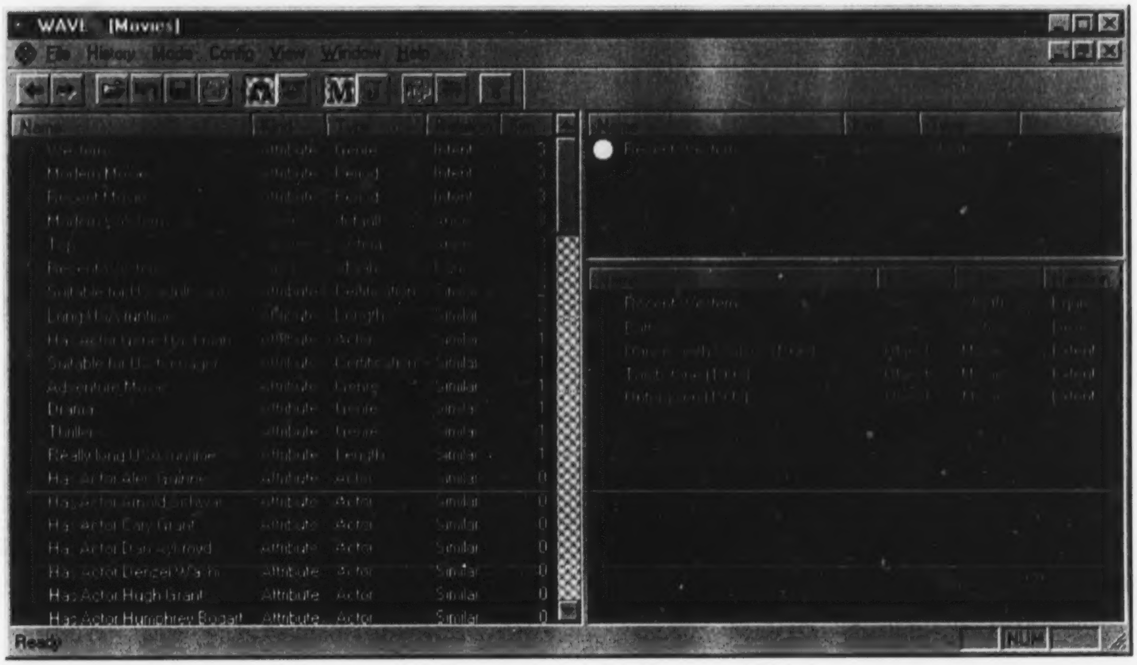


Figure 2: The WAVE Conceptual Browser Interface

References

- [Guha 1996] Guha, R.V. (1996). **Meta-Content Framework**.

- [Kent and Neuss, 1996] Kent, R.E., & Neuss, C. (1996). Conceptual Analysis of Hypertext. In *Advances in Hypertext for the World Wide Web*. Charles Nicholas and Jim Mayfield, editors. Springer-Verlag. Lecture Notes in Computer Science series.

- [Kent and Neuss, 1996] Kent, R.E., & Neuss, C. (1996). **Web Conceptual Space**. In *Proceedings of WebNet'96, the World Conference of the Web Society*, 15-19 October 1996, in San Francisco.

- [Luke et al, 1996] Luke, S., Spector, L., & Rager, D. (1996). **Ontology-Based Knowledge Discovery on the World-Wide Web**. In *Proceedings of the Workshop on Internet-based Information Systems*, AAAI-96 (Portland, Oregon).

- [Kent and Bowman, 1995] Kent, R.E., & Bowman, C.M. (1995). **Digital Libraries, Conceptual Knowledge Systems, and the Nebula Interface**. This white paper describes a formal conceptual basis for view-based organizations such as Nebula (the Nebula File System), SynFS (the Synopsis File System) and WAVE.

- [Wille 1982] Wille, R. (1982). Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts. In *Ordered Sets*, I. Rival, editor, pages 445-470. Reidel, Dordrecht-Boston.

BEST COPY AVAILABLE

Cineast - An Extensible Web Browser

Eckhart Köppen
Information Systems and Software Techniques
University of Essen, Germany
Eckhart.Koeppen@uni-essen.de

Gustaf Neumann
Information Systems and Software Techniques
University of Essen, Germany
neumann@wi-inf.uni-essen.de

Stefan Nusser
Management Information Systems
Vienna University of Economics and Business Administration, Austria
nusser@wu-wien.ac.at

Abstract: Cineast is a freely available, extensible Web browser which intends to provide an environment for prototyping new client side Internet technologies. Cineast has built-in support for HTML 3.2, fill-out forms, tables and incremental loading of documents. The browser itself uses the interpreted Wafe environment for implementing the high level control structures. The basic functionality is integrated into the Wafe package and coded in C, which means that the browser gains performance from the speed of compiled code while main aspects of the application can still be changed without recompilation. The network functionality is provided through the integration of the W3C Reference Library. The presentation of HTML documents is handled by the new Kino widget class which provides a flexible and extensible mechanism for parsing and rendering SGML like languages.

1 Introduction

Current development efforts in the domain of W3-applications concentrate on server-side enhancements. The reason for this is the well defined CGI interface, which encourages developers to enhance the server's capabilities. The development of new features on the client side is dominated by a few companies, which have the possibility to integrate new concepts in their browser products. So far, the only way of extending browsers are helper-applications (offering a poor integration with the browser) and plug-ins (being of highly vendor-specific nature). There are many enhancements such as access to new protocols, HTML extensions or peer-to-peer communication, which are impossible to realize this way. As a consequence, the development of such features in a Web browsing environment are mainly in the hand of two or three companies.

As a solution to this problem, we propose our concept of an extensible Web browser called Cineast. It is freely available and can be used as a prototyping environment for new Internet technologies. We achieve this flexibility by several means:

- We use compiled code for providing the basic functionality, but rely on the high-level Wafe [Neumann and Nusser (1993)] environment to implement the main features of the browser. This concept is comparable to that of a 4th generation language. The Wafe environment combines Tcl [Ousterhout (1990)] with MIT's X Toolkit [McCormack et al. (1990)] (Xt) and provides easy means to integrate further packages. Performance critical operations, as for example protocol implementations, HTML parsing or rendering are implemented in C. High level functionality, as user-interface semantics or network request coordination is implemented at the interpreter level and can therefore be modified without recompilation. This feature makes Cineast also an ideal platform for experimenting with e.g. mobile code systems.
- W3C's libwww [Frystyk-Nielsen et al. (1997)] serves as the basis for our network protocol functionality. The integration of libwww into the Wafe package gives us access to the functionality of this library, which is itself of very extensible nature. For example, new protocol suites, MIME-types or transfer encodings can be added in a straightforward manner. There is an ongoing effort of the W3C as well as of other people to keep the library up to date with current developments. The W3C itself is using libwww as a vehicle for

testing protocol extensions (such as PEP [Connolly et al. (1997)]), which made it the logical choice for our project. We have added HTTP over SSL support to Cineast by making use of libwww's protocol extension mechanism. The networking and security features mentioned here are described in more detail in an other paper [Neumann and Nusser 1997].

- For presentation purposes we use a widget class called Kino [Köppen (1996)], which was especially developed with the goal of flexibility and extensibility in mind. The Kino widget class contains a parser for SGML like languages and a rendering engine that is capable of managing arbitrary child widgets (called insets) which can be created in response to unknown tags encountered in the HTML source. Our support of the HTML FORM or IMG tags is completely based on this mechanism. The Cineast browser supports HTML 3.2.

In addition to this, Cineast supports a list of advanced features such as:

- multiple browser instances,
- full support of incremental loading and display (even incremental TABLEs),
- multiple simultaneous requests,
- request folding,
- request-wise transfer monitor,
- scroll linking of HTML source text and rendered display,
- built-in support for GIF, JPEG, PNG, XPM and XBM images.

The two main building blocks (libwww and Kino) will be presented in brief below before the discussion of Cineast itself.

2 The W3C Reference library

Figure 1 shows the main components of the W3C Reference Library and their interactions. A more detailed description can be found in the library's documentation [Frystyk-Nielsen et al. (1997)] or in the paper presented at the Fifth International WWW-conference [Frystyk-Nielsen (1997)]. The *Protocol Manager* is used to coordinate network access for application level protocols. Note that the protocol modules shown in Figure 1 are not part of the library core, although these are the modules, which are shipped with the current version of libwww (version 5.0a). The *Protocol Manager* furthermore provides functions for registering new protocols. We made use of the library's protocol extension mechanism by implementing HTTP over SSL [Netscape Corp. (1996)]. This gives the Cineast browser access to state-of-the-art security technology and allows us to experiment with new Internet security concepts.

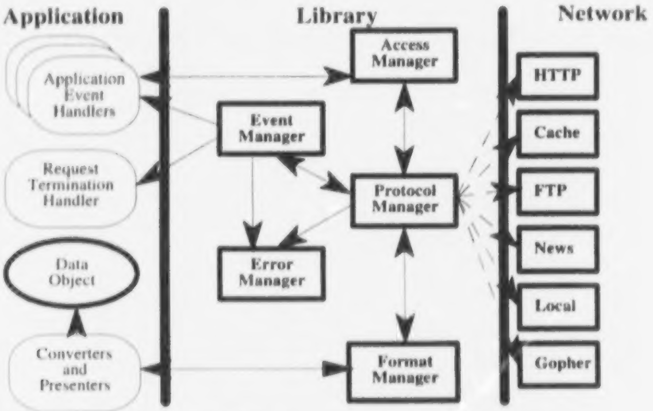


Figure 1: W3C Reference Library Architecture

The *Access Manager* is the main entry-point for applications into libwww's functionality. It comprises several functions for downloading and uploading URLs. Any error messages and warnings which arise during this process are collected by the *Error Manager* and can than be accessed by the application.

BEST COPY AVAILABLE

The *Format Manager* takes care of any conversions of the incoming or outgoing data. It will handle content-encodings or character-set conversions, as well as the final presentation of a downloaded object to the user, deploying previously registered converters and presenters.

Although the W3C Reference Library comes with its own *Event Manager*, this module is not part of the so called library core. It is the *Event Manager's* responsibility to trigger the protocol modules of libwww whenever data can be read or written from or to the network. This might in some cases dispatch some of the previously registered application-level event handlers, as for example the request termination handler which notifies the application of the request's completion. In our implementation, we actually use the event handling mechanism of Xt which allows us to integrate the handling of network and GUI events. This is of crucial importance for the response behavior of any network application - the same event handling mechanism should be in charge for dispatching user events and network events.

3 The Kino Widget Class

The Kino widget class is an Xt widget class written in C. It implements parsing, formatting and rendering of HTML text. But unlike other tools, it is easily extendible through the Xt callback mechanism. Though the parser of the W3 consortium's libwww and other parsers use this mechanism as well, the Kino widget class goes further by letting the application programmer control most of the internals of the widget. Among these internals are for example the layout information, the HTML source text and much more details. One of the most powerful features is the ability to add insets to the HTML text. These insets can be any kind of widget, even another Kino widget.

The Kino widget has to fulfill three major tasks like any other HTML displaying tool: parse the HTML source text, arrange the parsed elements of the source text and display the elements. Furthermore, proper handling of incremental source text completion is an important feature. Beside these points, the extendibility of the Kino widget requires more functionality:

- provide a clean model for accessing the Kino widget's internals
- interaction with other widgets
- provide a uniform interface for adding custom extensions to the Kino widget

This functionality is realized by three sub-objects: a parser, a layouter and a painter. These objects work on a set of data objects, mainly a list of parsed source text elements (called PData objects) and a list of layouted lines (called Lines structure). Figure 2 shows the interaction of the objects.

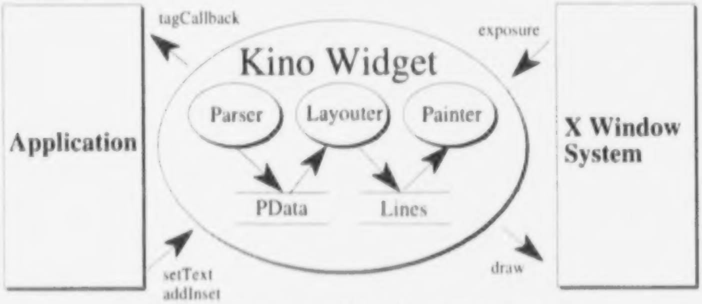


Figure 2: Overview of the Kino Widget Class

The parser is responsible for breaking up the source text into words and tags, which are the only recognized elements. It builds a list of parsed text elements made up of PData objects (these will be discussed further down). Any extension of the core Kino widget class can insert elements into this list during the parsing process such as simple words or more complex style data. The Kino widget itself just adds parsed words to the list.

After the parser (and the Kino extensions) have constructed the PData list, the layouter arranges the elements into displayable lines using the Lines object. The layout of the HTML text is constrained by the available width and the default text style. The layout process is triggered whenever the available width or the default style changes. Since these conditions occur quite often, the layout process has to be more optimized than the parsing process. The layouter itself optimizes the Lines structure for the painter by calculating as much position data as possible. The painter handles mostly exposure events from the window system but is also used internally for translating screen coordinates to PData elements and source text positions.

The PData objects are the building blocks of the parsed text. The most important elements are words, style and alignment data, table data and insets. These elements can be added when a tag is handled. The parser offers a programmatic interface for the PData list as well as two stacks used for nesting style and alignment data. If the application adds an inset to the PData list, it will be displayed at the current position on the line or aligned to the left or right margins of the text. The Lines structure contains the relation between the PData elements and the corresponding screen positions. It is used by the painter to update the display or to translate screen coordinates to source text positions.

The parsing process is the first point where the extendibility of the Kino widget is implemented: whenever a tag is encountered, the *tag* callback (a resource of the Kino widget with the name `tagCallback`) is invoked. The Kino widget itself does not process the tags further, so the task of handling the tags appropriately is up to the Kino extensions. These extensions can register a callback function for the tag callback using standard Xt functions, which makes the core Kino widget quite simple. The tag's attributes and their values are passed as a parameter to the callback functions.

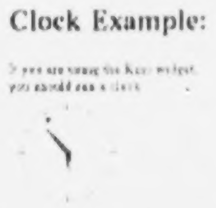
The standard Kino extensions mostly add text or style data. But by adding insets to the PData list with the `XkAddInset` command, more complex compound documents can be constructed. To demonstrate this feature the Kino widget is extended to handle the `CLOCK` tag from Tcl:

```
proc handleTag {w tag atts} {
  switch -exact $tag {
    CLOCK {
      XkAddInset $w [Clock c $w \
      width 100 height 100 \
      update 1 background pink] bottom
    }
}
```

This tag handler adds a Clock widget whenever the tag `<CLOCK>` appears in the source text. A text like

```
<H1>Clock Example:</H1>
If you are using the Kino widget, you should see a clock
<CLOCK>
```

produces output as seen in Figure 3 where the Clock widget displays the current time and updates itself every second.



Clock Example:

Figure 3: A Clock Widget as an Inset

Another feature of the Kino widget is its ability to change the HTML source text "on the fly", e.g. the Kino widget lets the application programmer change the text after the current parsing position (tag rewriting). By this means it is easy to implement a configurable filter that produces different HTML documents depending on a style guide. Another possible scenario is a client-side interface that allows any script to insert (and change) the source text, e.g. as a result of a database query, or one can handle semantic tags this way.

BEST COPY AVAILABLE

A semantic mark up like

```
<PERSON>Gustaf Neumann</PERSON>
<AFFILIATION>University of Essen, Germany</AFFILIATION>
```

can be processed can result in an appearance like

Gustaf Neumann is a Person and works for *University of Essen, Germany*

by defining the tag procedure in the browser like

```
proc tag {w tag atts} {
  switch -exact $tag {
    AFFILIATION  { XkChangeCurrentText $w "and works for <I>" 0 }
    /AFFILIATION { XkChangeCurrentText $w "</I>" 0 }
    PERSON       { XkChangeCurrentText $w "<B>" 0 }
    /PERSON      { XkChangeCurrentText $w "</B> is a Person " 0 }
  }
}
```

4 The Cineast Browser

For better code reuse we decided to use OTcl [Wetherall and Lindblad (1995)] rather than Tcl as the base implementation language of the browser. Several classes are used to implement the functionality. A `RequestHandler` handles the life cycle of a request, it has sub-classes for requests for HTML texts and images. Since images are implemented based on the inset capability of the Kino widget class, `Image` inherits from both `RequestHandler` (in order to control the transfer of the file) and `Widget` (to display the image).

The `RequestManager` class keeps track which requests are active per browser instance and aborts requests if necessary. The `HistoryManager` class handles the history of URLs for the handling of the Back and Forward buttons as well per browser instance. Finally the dialog classes are for `mailto:` tags (`MailDialog`), for HTML source browsing and editing (`EditDialog`) and for the transfer monitor (`TransferDialog`) which display transfer statistics on a per request basis and allows termination of single requests. A screenshot of the Browser is shown in Figure 4.

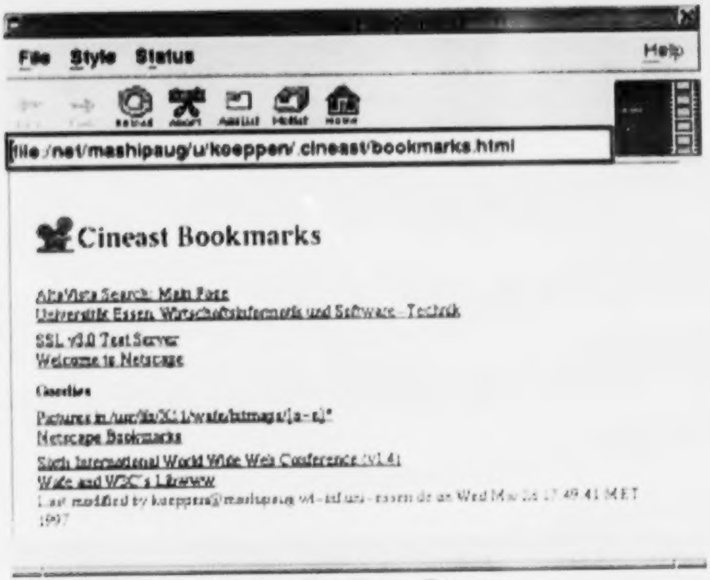
5 Conclusions and Future Work

With Cineast, we present a flexible web browser which is implemented in OTcl and built on top of the Wafe environment. We use libwww for networking functionality and a highly flexible widget class named Kino for HTML rendering. Our basic theme is that we try to combine and to configure efficiently implemented library functions (typically in C) in an as flexible as possible way using Tcl. The flexibility of Tcl (and OTcl) allows to reduce the development time for the sometimes elaborate configurations of the used components and to concentrate on the application tasks. We believe that our environment is one of the most powerful and flexible implementation environments for Web client development currently available. It is straightforward to extend it for:

- electronic commerce (experiment with various electronic payment approaches),
- non-standard Web client extensions (such as mobile code, peer-to-peer document exchange),
- for the development for embedded or specialized browsers (e.g. for certain application domains) and improving Web accessability for user groups such as handicapped persons, or as a
- platform for an Intranet development environment (supporting enabled forms, applets, database access, integration of push-model (email) and the pull model (WWW)), etc.

Our environment incorporates the basic security infrastructure necessary for such projects together with non-standard techniques (such as full tag handling and tag rewriting or insets) which provide more flexibility than plug-ins can offer. Future work includes the use of style sheets like CSS and the implementation of XML to further enhance flexiblity.

BEST COPY AVAILABLE



file:/net/mashipaug/u/koeppen/.cineast/bookmarks.html

Cineast Bookmarks

AltaVista Search: Main Page
Universität Essen, Wirtschaftsinformatik und Software - Technik

SSL v3.0 Test Server
Welcome to Netscape

Pictures in /usr/lib/X11/wafe/bitmaps/[a-z]*
Netscape Bookmarks

Sixth International World Wide Web Conference (v1.4)
Wafe and W3C's Libwww

Figure 4: The Cineast Browser

6 References

[Connolly et al. (1997)] Dan Connolly, Rohit Khare, Henrik Frystyk-Nielsen: *PEP: An extension mechanism for HTTP*, work in progress. See http://www.w3.org/TR/WD-http-pep.html in the W3C technical reports section (http://www.w3.org/TR/), July 1997.

[Frystyk-Nielsen et al. 1997] Henrik Frystyk-Nielsen: *Libwww - The W3CReference Library*, http://www.w3.org/Library/, April 1997.

[Frystyk-Nielsen (1997)] Henrik Frystyk-Nielsen: *W3C Reference Library*, position paper for the workshop: "Programming the Web - a search for APIs" at the Fifth International WWW-Conference (http://www.cs.vu.nl/~eliens/WWW5/papers/W3C.html), April 1997.

[Köppen (1996)] Eckhart Köppen: *Entwicklung eines erweiterbaren Widgets zur Anzeige von HTML-Texten*, Master's Thesis, University of Essen, Germany, 1996.

[McCormack et al. (1990)] Joel McCormack, Paul Asente, Ralph Swick: *X Toolkit Intrinsics-C Language Interface*, Massachusetts Institute of Technology and Digital Equipment Corporation, 1990.

[Netscape Corp. (1996)] Netscape Communications Corporation: *The SSL Protocol*, http://home.netscape.com/eng/ssl3/ssl-toc.html, March 1996.

[Neumann and Nusser (1993)] Gustaf Neumann, Stefan Nusser: *Wafe - An X Toolkit Based Frontend for Application Programs in Various Programming Languages*, USENIX Winter 1993 Technical Conference, San Diego, California, January 25-29, 1993.

[Neumann and Nusser (1997)] Gustaf Neumann, Stefan Nusser: *A Framework and Prototyping Environment for a W3 Security Architecture*, Proceedings of CMS'97, Chapman & Hall, September 1997.

[Ousterhout (1990)] John K. Ousterhout: *Tcl: An embeddable Command Language*, Proceeding USENIX Winter Conference, January 1990.

[Wetherall and Lindblad (1995)] David Wetherall and Christopher J. Lindblad: *Extending Tcl for Dynamic Object-Oriented Programming*, Proceedings of the Tcl/Tk Workshop '95, Toronto, July 1995.

BEST COPY AVAILABLE

Internet Based Collaboration: the Presence and Perspectives

Miloš Kravėi k
Faculty of Mathematics and Physics, Comenius University, Slovakia
E-mail: kravcik@fmph.uniba.sk

Abstract: Collaboration is a complex process and current information and communication technologies provide various facilities to support it. Internet services have strong impact on the quality of remote collaboration but their potential has definitely not been exhausted in this field.

This paper discusses the current opportunities the Internet provides for collaborative work focused on creation of a tangible artifact, especially when expository writing is concerned. Cognitive models and design analysis can help to reveal restrictions and outline perspectives offered by current information technology for collaboration support. We compare first and second generation of distributed hypermedia systems from this point of view.

1 Introduction

The history of computer supported collaboration is originated in the idea of intelligence amplification [Bush 45]. Later efforts in augmenting human intellect [Engelbart 63] led to the integrative paradigm: Concurrent Development, Integration and Application of Knowledge (CoDIAK). After Engelbart's Augment/NLS, mainly in the last decade, several collaboration supporting systems have been developed (for instance SEPIA, ABC, HB1,...), which brought various models of collaboration and enhanced our knowledge in this area. Nevertheless these are research systems and thus not widely used. The other approach how to meet the collaborative needs of the Internet users is to further develop and integrate popular tools and systems.

Recently the term *open collaboration* has become used more and more often, especially by Netscape Communications. It is set up on the Internet infrastructure and should lead from monolithic, proprietary architectures to a single architecture based on open Internet standards. The main objective of this trend is interoperability together with a rich feature set on an open platform.

The paper consists of two main parts. What follows is an overview of current communication and collaboration opportunities offered by the Internet and an outline of possible improvements. A special interest is devoted to the authoring process in the next section. First we mention several cognitive models for writing. These are followed by issues related to the design process together with the potential of second generation hypermedia systems [Maurer 96] to solve them in comparison with the first generation. Second generation hypermedia systems (HyperWave) are extensions of first generation systems (World-Wide Web). The second generation includes features relevant for collaboration support, for example bi-directional links, separate link database, hierarchical structure, document attributes, integrated search facilities, and version control.

2 Internet Services

Up to now the main impact of the Internet can be seen in the areas of communication and publishing. This has direct influence also on the field of collaboration, but it seems the Internet still does not fully support collaboration in general. In the following we try to explain this opinion.

2.1 Communication

Communication is a crucial part of collaboration and the Internet has changed it essentially from its beginning. It was electronic mail which attracted lots of people to the Internet. E-mail increased efficiency of communication by overcoming both temporal and spatial barriers. Additionally the leadership roles could be spread more evenly among participants. Besides asynchronous forms of communication also synchronous ones (in real time) have been deployed on the Internet, using text, audio, and video channels, as well as white boards.

But there can also be other channels for synchronous communication. Hypermedia has been discovered as a natural means for collaboration support (see [Engelbart 95], [Streitz 93]). The Web is now a huge repository of information and synchronously communicating users often need to use Web documents for demonstrative or argumentation purposes. A hypermedia channel enhances communication possibilities enabling collective browsing with shared documents. Up to now this way of communication is not so common as those mentioned above. Netscape Communicator is the first widely used program supporting shared browsing of documents between two users.

Another approach [Kravcik & Mederly 97], which is more general from the number of participants point of view, uses two types of viewers. A user has a private viewer for private browsing and a group viewer for sharing documents with other participants. He or she can select any document from the private viewer for its later sharing and redistribute it when needed. This architecture is based on a TCP/IP daemon for inter-application communication [Mederly et al. 97] and realised using widely used Web browsers (Netscape Navigator) and servers. In practice it is used together with a multiplatform videoconferencing system (CU-SeeMe).

Collaborative browsing is also the main topic of the *CoBrow* project [Sidler et al. 97] which aims to extend the current World-Wide Web by the concept of *meeting places*. Based on the Internet standards meeting places should enable applications like online meetings, help desks and forums.

2.2 Collaboration

Communication is a necessary part of collaboration and currently most of the Internet users cooperate just by exchanging information, e.g. working versions of documents. But if the objective of collaboration is a complex tangible artifact there is a need to employ more sophisticated models supporting gradual development of such a product by a group of authors. While the Internet has changed essentially communication and publishing which concern a part of the collaboration process and its outcomes, collaboration itself is still waiting for similar improvements.

In addition to communication mentioned in the previous section collaboration consists of individual and collective creation of an artifact. We can again distinguish its asynchronous and synchronous forms. In the first case individuals create parts of the artifact and the colleagues express their opinions or modify it. The other possibility means simultaneous collective development of the artifact. Considering the Internet potential in the area of collaboration we are primarily focusing on the development of documents (this general term includes also hyperdocuments, i.e. documents with non-linear or multi-dimensional structure). A natural medium for this purpose is hypermedia as a means to express and represent inter-object relationships and (alternative) structure(s) of these artifacts.

Gradual iterative development of content and structure is the essence of asynchronous collaboration. Actually it is very closely related to the above mentioned asynchronous communication when (parts of) documents and comments are exchanged. This relationship between asynchronous communication and collaboration is manifested not only by the expected single standard (HTML) for all user generated content [Udell 97]. Distributed hypermedia systems should also allow to open a discussion on an arbitrary Web document as well as to be notified when a new document (it means also a comment) appears in a specified Web area. Thus a user could add a comment directly to the hypermedia structure and anybody who is interested (and has access permissions) would be immediately notified. Using an alternative approach a user responds to a received message containing a (part of a) document or a comment and his/her respond is automatically added to the non-linear structure. Hypermedia serve as a means to archive structured discussions and e-mail informs users of new submissions in them.

Synchronous collaboration in the sense of tangible artifact development is mostly known as shared editing of a (text) document. But shared viewing and modification of hypermedia structure is not so common at all.

3 Authoring

If we want to further investigate collaborative design and development of documents we need deeper understanding of the authoring process. In the following we describe several cognitive models of writing, which is the most typical authoring process. Then we discuss design issues concerning especially hypermedia applications. Suggestions how to apply the mentioned principles in practice are outlined too.

3.1. Cognitive Models for Writing

Our brief overview of cognitive models for writing begins with the most simple one of them [Rada 91]. Rada assumes that the final form of a text is determined by goal and audience. According to this model writing consists of three phases:

- exploring (creation of unstructured notes)
- organising (hierarchical ordering of the notes)
- encoding (writing the target document)

The model by Hayes and Flower (see [Hayes & Flower 80], [Flower & Hayes 84]) has three basic components:

- task environment (formed by a writing assignment and a preliminary text)
- writer's long term memory (including knowledge of the topic and of the audience as well as writing plans)
- processor (containing three high-level processes and a monitor, providing overall control of the writing system)

Those three high level processes of the processor are:

- planning (generating ideas, organising document scheme, and goal-setting to control movement among these two subprocesses)
- translating (encoding the ideas into continuous prose)
- reviewing (reading the produced text and its editing)

Smith [Smith 94] describes a framework based on a set of cognitive modes which are used by individuals to perform a task, and on strategies exploited when moving among these modes. A cognitive mode is considered as a particular way of thinking used for a particular purpose. Seven cognitive modes are distinguished:

- exploration
- situational analysis (analysing objectives and audiences, prioritising)
- organising
- writing
- editing - global organisation
- editing - coherence relations (between sentences and paragraphs)
- editing - expression (linguistic analysis)

We can see that these cognitive models differ just in levels of abstraction or degrees of detail they employ.

As individual knowledge and skills are restricted and the time factor often plays an important role, it makes sense to consider also a cognitive model of collaborative writing. Such a model [Sefranek & Kravcik 97] corresponds to those mentioned above. It consists of a knowledge base and a text base which can be seen as the writer's memory and the task environment respectively, if we use the Hayes' and Flower's terminology. The processor oscillates between development of a document content and structure, using functions as generating, elaborating, and reviewing. The writer's ability to spontaneously restructure his or her knowledge in adaptive response to changes in situational demands is crucial. This ability is known as cognitive flexibility [Spiro & Jehng 90] and plays a key role in this model, which includes functions like focusing attention, changing the level of abstraction and detail, as well as distinguishing relevant information in the knowledge base. The knowledge base is modelled as a heterogeneous multi-layered semantic network. The processes of reading and writing are composed of the operations on the network. Collaborative writing is explained over a structured and evolving complex of private and common subnetworks.

From the implementation point of view we can find several essential features in second generation hypermedia systems which are missing in the first generation. Hierarchical structure is a natural means to model levels of abstraction. Separation of the link database from the documents content enables creation of alternative structures over the same documents. Together with version control this helps a lot during gradual elaboration of documents structure and content. Structured discussions are crucial in the reviewing phase. They are well supported by bi-directional and typed links which enable visualisation of the structure. Cognitive flexibility aspects depend on sophisticated searching and filtering that can operate with document attributes in second generation hypermedia systems.

3.2 Design

The results of collaborative work are usually not simple documents, but rather complex artifacts. As we have already mentioned hypermedia is a natural means for collaboration, even when the outcome is to be a linear document. Thus hypermedia can be seen both as content of and medium for collaborative work [Streitz 93].

In the development of complex artifacts the quality of the design process is essential. Taking into account the role of hypermedia in collaborative work we are interested in the hypermedia design process. This area has been investigated a lot recently.

3.2.1 Design Process

Hypermedia design can be considered as a "Brownian motion" in a two-dimensional space [Nanard & Nanard 95] where one dimension is formed by formal techniques (concepts elicitation, navigation model, abstract interface, implementation model, testing) and the other are mental processes as mentioned above (generating material, organising and structuring, reorganising and updating, evaluating). While the formal design technique is determined by the concrete application, mental processes are more universal.

The design process is both top-down and bottom-up. In the first case from abstractions at the conceptual level (knowledge base) instances at the implementation level are derived. Using the opposite approach a set of instances is conceptualised into a generic structure. Semantics of elements at the conceptual level can be captured by abstract semantic types. These can be represented in the form of semantic networks.

In the second generation hypermedia systems metainformation stored in document attributes can be used for typing. As representation of semantic networks is similar to representation of hypermedia ones abstractions can be modelled by means of hypermedia systems. Designers need a suitable way to handle structure of both hypertext and semantic networks so a graphical structure editor would be helpful for them. It is also necessary to keep consistency at the conceptual level as well as relationships between abstractions and instances.

3.2.2 Design Issues

Another critical aspect of hypermedia design is comprehension of hyperdocuments [Thuering et al. 95]. This is mostly influenced by two factors: coherence and cognitive overhead, in the first case positively and in the second one negatively.

Coherence is determined by reader's ability to construct a mental model corresponding to a possible world. "Small scale" relations (between clauses or sentences) establish local coherence. "Large scale" connections (conclusions drawn from several clauses, sentences, paragraphs, chapters) establish global coherence of a text. Considering hyperdocuments both local and global coherence should be taken into account at two levels - the node level (within nodes) and the net level (between nodes).

Cognitive overhead is the additional effort necessary to maintain several tasks at one time when reading a hyperdocument [Conklin 87]. This includes orientation, navigation, and user-interface adjustment.

Attempts to increase coherence and reduce cognitive overhead for better comprehension imply cognitive design issues for creating hyperdocuments. In response to these issues eight design principles have been proposed in [Thuering et al. 95]:

1. *Typed link labels* to represent semantic relations between information units
2. *Indicating equivalencies between information units* to reduce the impression of fragmentation
3. *Preserving the context of information units* to reduce the impression of fragmentation
4. *Higher-order information units* to structure the document
5. *Visualising the structure of the document* to provide an overview of the hyperdocument
6. *Including cues into the visualised structure for the reader's current position* to improve orientation
7. *Navigation facilities which cover aspects of direction and distance* to facilitate navigation
8. *Stable screen layout* to reduce the effort for interface adjustment

We have already mentioned the possibility to type documents by means of document attributes in second generation hypermedia systems. HyperWave enables also typing of links. The attributes of a link including the link type are stored with its source anchor. To support local coherence the types should be visually indicated and together with the currently activated node also its predecessor should be kept on the screen. Both should also be indicated in the overall hierarchy, history, and a local map. These are very efficient second generation facilities to support global coherence as well as orientation and navigation. It would be good if also a net of nodes (possibly without a real content, representing just a concept instead of a real document) and links indicating semantic relations among them could be displayed. Assistance in searching for relationships between nodes would be helpful too. Stable screen layout reduces additional effort in user-interface adjustment.

4 Conclusion

Development of information technologies tends to better support for collaboration. The Internet provides an excellent infrastructure and efficient services in this respect. However there are problems that have not been addressed yet in commonly used systems. Our intention was to highlight some of them (a new communication channel, shared development of hypermedia structure, implementation of a cognitive model for collaborative writing, design issues) and to outline possible solutions. Further development of second generation distributed hypermedia systems provides promising perspectives for collaboration in the future.

5 References

[Bush 45] Bush, V. (1945). As we may think. *Atlantic Monthly*, 171(1), 101-108.

[Conklin 87] Conklin, J. (1987). Hypertext: An Introduction and Survey. *IEEE Computer*, 20(9), 17-41.

[Engelbart 63] Engelbart, D. C. (1963). A conceptual framework for the augmentation of man's intellect. In P. D. Howerton, D. C. Weeks, editors, *Vistas in Information Handling*, Vol. 1, Spartan Books, Washington D.C., 1-29.

[Engelbart 95] Engelbart, D. C. (1995). Toward Augmenting the Human Intellect and Boosting our Collective IQ. *Communications of the ACM*, 38(8), 30-33.

[Flower & Hayes 84] Flower, L. S., Hayes, J. R. (1984). Images, plans, and prose: The representation of meaning in writing. *Written Communication*, 1 (1984), 120-160.

[Hayes & Flower 80] Hayes, J. R., Flower, L. S. (1980). Identifying the organisation of writing processes. In L. W. Gregg, E. R. Steinberg (Eds.), *Cognitive processes in writing*. Hillsdale, NJ: Lawrence Erlbaum Associates, 3-30.

[Kravcik & Mederly 97] Kravèi k, M., Mederly, P. (1997). Toward Open Hypermedia Systems. *MIPRO International Conference on Hypermedia and Multimedia Systems*, Opatija, Croatia. 84-87.

[Maurer 96] Maurer, H. (1996). *HyperWave: The next generation Web solution*. Addison-Wesley.

[Mederly et al. 97] Mederly, P., Kravèi k, M., Bebják, A., Hine, N. A., Gordon, I. A. S., Beattie, W., Arnott, J. L., Arato, A., Seiler, F. P. (1997). A TCP/IP Daemon that allows inter-application communication across the Internet. *ED-MEDIA & ED-TELECOM 97*, Calgary, Canada.

[Nanard & Nanard 95] Nanard, J., Nanard, M. (1995). Hypertext Design Environments and the Hypertext Design Process. *Communications of the ACM*, 38(8), 49-56.

[Rada 91] Rada, R. (1991). *Hypertext: from text to expertext*. McGraw-Hill.

[Sefranek & Kravcik 97] Šefránek, J., Kravèi k, M. (1997). A model of collaborative writing. *Cognitive Systems*, 4(3-4), 401-428.

[Sidler et al. 97] Sidler, G., Scott, A., Wolf, H. (1997) Collaborative Browsing in the World-Wide Web. *8th Joint European Networking Conference, 1997*, Edinburgh, Scotland. (122-1)-(122-8).

[Smith 94] Smith, J. B. (1994). *Collective intelligence in computer based collaboration*. Lawrence Erlbaum Associates.

[Spiro & Jehng 90] Spiro, R. J., Jehng, J.-Ch. (1990) Cognitive Flexibility and Hypertext: Theory and Technology for the Nonlinear and Mutlidimensional Traversal of Complex Subject Matter. In D. Nix, R. Spiro (Eds.), *Cognition, Education, and Multimedia: Exploring Ideas in High Technology*. Hillsdale, NJ: Lawrence Erlbaum Associates. 163-204.

[Streitz 93] Streitz, N. (1993). The Role of Hypermedia for CSCW Applications. *Third European Conference on Computer Supported Cooperative Work*, September 1993, Milano, Italy.

[Thuering et al. 95] Thuering, M., Hannemann, J., Haake, J. M. (1995). Hypermedia and Cognition: Designing for Comprehension. *Communications of the ACM*, 38(8), 57-66.

[Udell 97] Udell J. (1997). Net Applications: Will Netscape Set the Standard? *Byte*, March 1997, 66-84.

A Framework for WWW-based Learning with Flexible Navigational Guidance

Christian Langenbach, Freimut Bodendorf
Dpt. of Information Systems, University of Erlangen-Nuremberg, Germany

{Christian.Langenbach|Freimut.Bodendorf}@wi2.wiso.uni-erlangen.de

Abstract: The quality of WWW-based learning depends on several critical success factors. Especially, the course materials on the WWW should not represent a one to one transfer of written lecture notes. Added values like interaction and dialogue components, training modules, etc. should be provided. This paper introduces the approach of the multimedia WWW-based teachware packages taking these demands into consideration. Applications are based on a series of modular reusable components providing core functionalities - especially flexible navigational guidance - and serving as a framework for developing WWW-based teachware packages with any type of content.

1 Motivation

WWW-based learning can offer a variety of interesting advantages compared to traditional Computer Based Training (CBT). Among others these include the inherently effortless distribution of learning materials combined with the ease of promptly updating courses and the possibility of simply reusing existing lecture content in a new context as well as many possibilities of collaborative learning within the network. To fully utilize the conceivable potential a few constraints, specific to WWW-based learning, need to be taken into consideration.
At the University of Erlangen-Nuremberg, which is spread out over an area of about 500 km^2, a project to create a "virtual campus", providing a new type of network based learning experience is currently being run. Part of this undertaking is to create a supported, open environment for teaching, learning and co-operation, which provides students with multimedia WWW-based teachware packages on various subjects. This work, which is being pursued as part of a teleteaching/telelearning project [Bodendorf, Grebner & Langenbach 1997], focuses on critical success factors of WWW-based learning. WWW-based teachware packages are integrated in the curriculum of students at the Faculty of Economics and Social Sciences. They encourage a self motivated acquisition of basic knowledge, which is then further refined during face-to-face lectures, tutorials and seminars.

2 A Framework for WWW-based Teachware Packages

Multimedia WWW-based teachware packages are implemented entirely in HTML, Java, JavaScript and WWW-compatible media formats, with no CGI processes taking place. This provides the means for online use on the Inter- and Intranet but also allows for offline use via CD-ROM. The only prerequisite for the use of the teachware packages is a Java-compatible browser as a front end, with no additional applications or plug-ins which need to be installed and configured.
The teachware packages are based on a series of modular reusable components, which were developed in the first stages of the project to provide core functionalities and to serve as a framework for developing WWW-based teachware packages with any type of content. The learner can make use of this framework functionality in form of a *control panel* (see [Fig. 1]) implemented in Java, where different buttons are dynamically enabled or disabled depending on the context of the information being displayed.
The browser window is subdivided into two frames. The presentation of course materials in the main frame (*information frame*) is managed through the control panel situated in the *control frame*. In addition to navigational guides the framework offers a variety of other functionalities to the user via the control panel. The corresponding components providing these functionalities are outlined below.

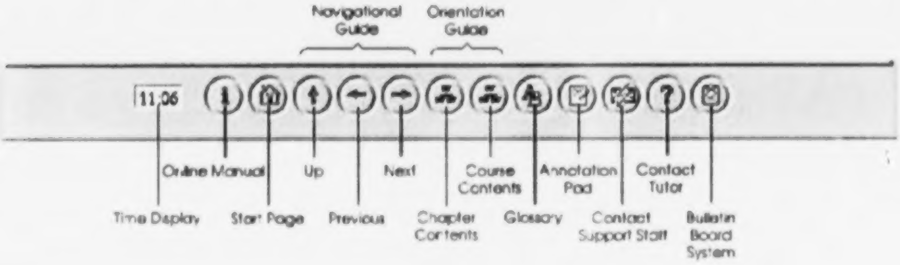

Navigational Guide
Orientation Guide
11:06
Online Manual
Up
Next
Course Contents
Annotation Pad
Contact Tutor
Time Display
Start Page
Previous
Chapter Contents
Glossary
Contact Support Staff
Bulletin Board System

Figure 1: Control Panel

2.1 Flexible Navigational Guides

Navigation through the course materials, which are arranged in a hierarchical way, is supported through the control panel Java applet with its three buttons *up*, *next* and *previous*, the latter offering a specific and context related functionality.

In principle the course content may be experienced in two ways: first through unstructured exploration, looking at table of contents, overview pages and by following links to more information within the teachware package (*user determined presentation*), or second by using the *next* button to follow a predefined guided tour step by step (*system determined presentation*). In the context of this second approach the *previous* button allows the user to go back one step on the guided tour, whereas the *up* button brings the user to the root of the hierarchically preceding course module. If the user initially decides to use the system determined presentation, but leaves the predefined path, for instance to follow links and cross-references to supplementary external resources on the WWW (*composite type presentation*), he or she can always return to the last viewed chapter of the online course by using the *previous* button.

[Fig. 2] shows a schematic representation of the functionality of the navigational guides within a composite type presentation.

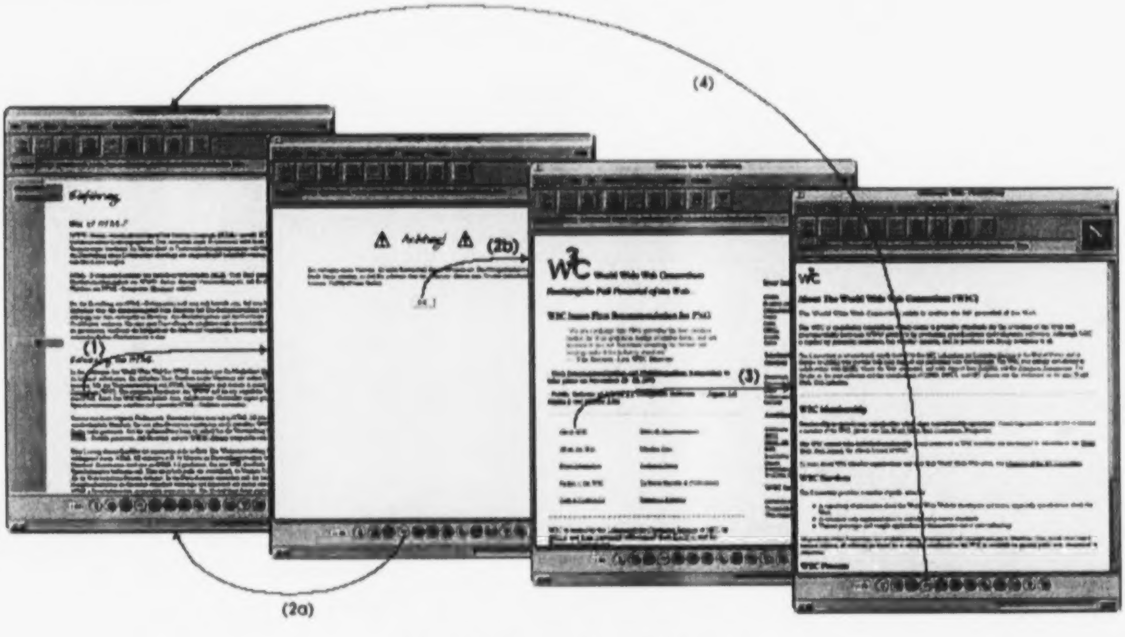


Figure 2: Navigational Guides

BEST COPY AVAILABLE

If the learner selects a link to an external source of information on the WWW, the system signals that he or she is about to leave the guided tour of the teachware package (1). At this point the user has the option to return to the guided tour by pressing the *previous* button (2a), or to follow the external link (2b), which is then loaded from the WWW into the information frame without relinquishing the functionality of the control panel which is preserved in the adjacent frame. Within the external sources of information on the WWW the user can follow any other links (3). By using the *previous* button he or she can always directly return to the last chapter within the guided tour of the teachware package (4), without having to use browser specific aides, such as multiple clicks on the "back" button.

This functionality is available in all situations, open to a departure from the predetermined guided tour, i. e. when looking up terms in the glossary or when consulting the online manual.

2.2 Orientation Guides

Within a multimedia WWW-based teachware package essentially two methods are used as an orientation guide. First a *colour coding* is applied to the course materials and second dynamically updated *tables of contents* for the individual chapter and the whole course can be switched on and off in a separate frame by using two corresponding buttons on the control panel. This approach enables the learner to always easily determine his or her exact position within the course without leaving the guided tour.

2.3 Interactive Components

In order to offer the learner interactive components as in standard progression tests a building block was developed in Java, which provides a simple definition of *free form and multiple choice questions and answers* to be integrated into the WWW-based teachware package.

Additionally the fundamental constructivistic requirement for *situation and context oriented learning* [Reinmann-Rothmeier & Mandl 1997] is fulfilled by providing interaction modules for specific areas of the course, where the learner can immediately apply the acquired knowledge to solve an authentic problem. For a HTML course for instance a JavaScript-based HTML test editor was developed, which allows the interactive creation and illustration of HTML documents, for example as a task within a case study, to fully apprehend the course material.

The interactive components are, in contrast to the other program modules introduced here, not accessible via the control panel, but through corresponding buttons within the course context of individual chapters.

2.4 Annotations

An *annotation pad* gives the learner the opportunity to take individual notes on each chapter, which can also be accessed out of context during a subsequent session with the corresponding teachware package.

2.5 Glossary

The *glossary*, which is subdivided into three levels, is always available to access definitions of unclear terms without having to go through any search procedures. Additional external resources from the WWW are easily linked if needed. For an in depth study of sources and literature in support of a specific area of interest, one could imagine an interface to an electronic library.

2.6 Online Manual

An integrated *online manual* is provided to help the user with regard to application specific questions, the functionality of the provided course modules as well as the available communication channels with the tutor, the technical support staff and other learners. By optionally opening another browser window, the online manual can be used in parallel to the teachware package.

2.7 Assistance by the Tutor and Technical Support Staff

During the online use of a teachware package in a distributed network based teaching and learning environment the learner must always be able to *contact a tutor and the supporting technical staff* through integrated media channels [Graesel, Bruhn, Mandl & Fischer 1996]. This allows the learner to receive help and to jointly find solutions to content related problems and technical as well as ergonomical difficulties. Some interfaces for this type of feedback and interaction have been already tied into the teachware package - reachable via the control panel - to use WWW-based synchronous and asynchronous communication tools, while the integration of others is anticipated. These mechanisms for interaction range from pre-addressed email forms over bulletin boards and whiteboard tools to shared application and videoconferencing systems.

2.8 Learner-Learner Communication

The *multilateral communication* among geographically distributed learners in an online course is also realized. One of the tools already provided is a multimedia WWW-based bulletin board system. In this forum for asynchronous discussion the tutor can create closed user groups, within which learners can jointly solve problems without the feeling of being observed by a third party. An additional feature of the system is its support for multimedia elements, for instance recorded audio contributions or video clips, which permits asynchronous communication not only on a textual level.

3 Reusability and Course Generation

The framework consisting of the program modules outlined above can be filled with multimedia course materials on any topic. At our university teaching materials which are produced within in a variety of other teleteaching and telelearning applications are frequently (re)used in the WWW-based teachware packages. These teaching materials consist of excerpts from lecture video recordings, corresponding digitized blackboard snapshots and overhead transparencies, as well as of contents from multimedia presentations, electronic lecture notes, excerpts from the course textbooks, exercises, etc. In addition, computer animations, textual, visual and audio components are added and external supplementary WWW resources are linked.

The integration of these elements does not require any changes to Java or JavaScript program code. Hence, the respective course author's tasks only include the subdivision of available electronic learning materials into suitable course modules and the integration of the materials into HTML documents, as well as linking the contents of the course to the supporting components outlined above, taking into account their specific advantages (i. e. navigational guidance, possibility of annotation, progression tests etc.) within the respective teachware package.

The guided tour within a teachware package can be laid out using two different methods. Both approaches described here do not rely on any CGI techniques and server interactions which are for instance used in [Goldberg, Salari & Swoboda 1996], [Kutschera 1996] and [Hauck 1996]. By determining the guided tour through a client-based Java/JavaScript approach one can realize a better performance. In addition, the teachware packages can be optionally distributed on CD-ROM and used offline since all the necessary information for the navigational guides can be accessed using the file protocol.

3.1 Integrated Guided Tour Definition

Within this approach, the initialization of the buttons *next*, *previous* and *up* in the control panel is achieved through a JavaScript instruction, individually on every HTML page. Each of these simple instructions merely contains information about the URL (Uniform Resource Locator) being assigned to each button for navigating back, forward or up within the guided tour and hence does not require any programming skills of the course author. When loading the document the JavaScript instructions are executed and through the Java-JavaScript communication the navigation buttons of the control panel Java applet are initialized by the corresponding method calls. If the learner in an online environment leaves the guided tour to freely explore external sources of information, the *previous* button is dynamically assigned the URL of the last page visited within the guided tour whereas the *next* and *up* buttons are being disabled.

The approach of the *integrated guided tour definition* described here can only be realized if the course author has access to the HTML documents that are to be displayed as part of the guided tour to add the corresponding initialization instructions to the pages. Additional external resources on the WWW can be *linked* and explored freely by the user leaving the guided tour, with the possibility to directly return to the last viewed chapter of the guided tour anytime using the *previous* button as described above. An integration of external WWW resources *into* the guided tour of the course is not feasible using this approach.

3.2 Separated Guided Tour Definition

An increased range of possibilities is created through the concept of a *separated guided tour definition*. Characteristic of this approach is the separation of HTML documents containing the course material and the meta information defining the guided tour. The course author hence is in a position to create an online course using any teaching materials available on the WWW, whether these are produced internally or externally, supported by the functionality of the navigational guides as described above. Only an common text editor is required for the course author to define the intended guided tour of the teachware package, by sequentially listing all URLs of the pages that are to be part of the course. The hierarchical layer *n* of the individual course modules within the planned teachware package is declared in this list, simply by prefixing a [L*n*]-tag to each URL. [Fig. 3] shows a simple example of a separated guided tour definition for a teachware package on "Internet and WWW".

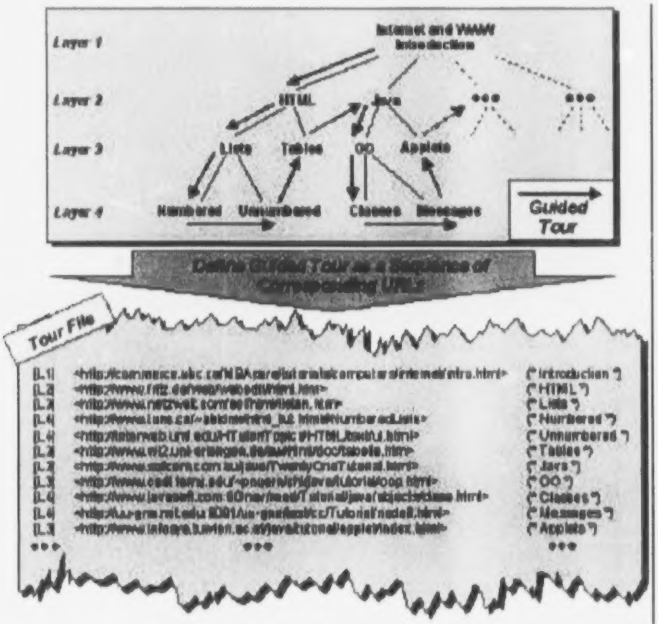

Layer 1
Layer 2
Layer 3
Layer 4
Internet and WWW Introduction
HTML
Java
Lists
Tables
OO
Applets
Numbered
Unnumbered
Classes
Messages
Guided Tour
Define Guided Tour as a Sequence of Corresponding URLs
Tour File
("Introduction")
("HTML")
("Lists")
("Numbered")
("Unnumbered")
("Tables")
("Java")
("OO")
("Classes")
("Messages")
("Applets")

Figure 3: Separated Guided Tour Definition

This URL list - stored in a so-called *tour file* - is used by the navigation applet as a meta information to dynamically initialize the navigational buttons in the control panel. The assignment process is as follows:

BEST COPY AVAILABLE

- ☐ While loading a document from the teachware package the tour file is searched for this actual URL.
- ☐ If an entry is found the *previous* button is assigned the URL of the line directly above this entry and the *next* button is given the value of the URL of the following line.
- ☐ In order to initialize the *up* button, the URL of the line with the next lowest *n* in the [L*n*]-tag is used.

Once the learner leaves the predefined guided tour to explore additional external information on the WWW the navigational applet will not be able to find this new URL on the list. In this case, analogous to the integrated guided tour definition, the *previous* button is assigned the last viewed URL within the guided tour of the teachware package and the *next* and *up* buttons are disabled (see [Integrated Tour Definition]).

The concept of a separated guided tour definition does not require any changes to the source of individual course chapters, in contrast to the integrated guided tour definition. This particularly facilitates the task of changing the guided tour by adding, removing or reorganising the course content. Since course materials and meta information are separated, these changes only need to be applied to the tour file described above. In addition, through this approach the course documents can be referenced and used in the tour files of any online courses. Furthermore, any available WWW resources can be integrated into the guided tour of a teachware package. Hence, the requirement of multiple reusability of existing WWW-based teaching modules can be fulfilled in a very easy way.

4 Conclusions and Outlook

The positive feedback to this prototype framework for multimedia WWW-based teachware packages - award winning in this year's software competition of the German Academic Software Co-operation - encourages an extended evaluation beyond the boundaries of our university. An evaluation of the packages is planned at several German and Austrian universities and colleges, as well as in companies and in the area of teacher training. The interest shown by companies and non-university institutions suggests that multimedia WWW-based teachware packages may not only offer a suitable medium for WWW-based learning in higher education, but also a means for knowledge transfer between universities and business.

The development of additional reusable components for multimedia WWW-based teachware packages is planned. Particular emphasis will be placed on the development of Java-based interactive components as well as on an extended navigational support with graphical overviews and an adaptive approach.

References

[Bodendorf, Grebner and Langenbach 1997] Bodendorf, F., Grebner, R., & Langenbach, C. (1997). Telelearning in the Virtual Lecture Theatre. Displays, 17 (3-4), 147-151.

[Goldberg, Salari and Swoboda 1996] Goldberg, M. W., Salari, S., & Swoboda, P. (1996). World Wide Web - Course Tool: An Environment for Building WWW-based Courses. Computer Networks and ISDN Systems, 28 (7-11), 1219-1231.

[Graesel, Bruhn, Mandl and Fischer 1996] Graesel, C., Bruhn, J., Mandl, H., & Fischer, F. (1996). Lernen mit Computernetzen aus konstruktivistischer Perspektive. Research Paper No. 73, Dpt. of Empirical Pedagogics and Pedagogical Psychology, Ludwig-Maximilians-University of Munich, Munich, 17-20.

[Hauck 1996] Hauck, F. J. (1996). Supporting Hierachical Guided Tours in the World Wide Web. Computer Networks and ISDN Systems, 28 (7-11), 1233-1242.

[Kutschera 1996] Kutschera, P. (1996). Combining Database Technology with the World Wide Web for Tele-teaching Environments. In Swiridow, A. P., Widmayer, P., Oberhoff, W.-D., & Unger, H. (Eds.). New Media for Education and Training in Computer Science. St. Augustin: Infix, 99-108.

[Reinmann-Rothmeier and Mandl 1997] Reinmann-Rothmeier, G., & Mandl, H. (1997). Multimediale Lernumgebungen. In Geissler, K. A., v. Landsberg, G., & Reinartz, M. (Eds.). Handbuch Personalentwicklung und Training. Deutscher Wirtschaftsdienst, Cologne (forthcoming).

Acknowledgements

The work is being pursued in the context of a teleteaching/telelearning project, funded by the Association for Promoting a German Research Network (DFN-Verein) and the German Federal Ministry of Education, Science, Research and Technology. The authors thank Michael Stadelmann and Malte Stoll for their excellent work in the implementation phase.

Interface Design Issues for Web-based Collaborative Learning Systems

Glenn E. Lautenbacher
Dirk E. Mahling

Department of Information Science and Telecommunications
University of Pittsburgh

Abstract: This paper presents interface guidelines from our research in developing a collaborative, problem-based learning environment for the WWW. The lessons learned are based on student use and evaluation of three interface prototypes over the course of three years spanning several domains. Insight into appropriate windowing strategies, choice of menu structure and presentation, menus as a group coordination mechanism, and group annotation mechanisms are discussed. Extensions to our interface based on these findings are discussed and directions for future research are given.

1.0 Introduction

Over the past few years, the WWW has evolved from being a simple, document delivery mechanism to an increasingly complex, dynamic, and interactive environment. The ubiquitous nature of the WWW and the platform-independence that it provides to systems developers, provides opportunities to deploy systems that can be accessed virtually anywhere in the world, at any time, and on a wide variety of computers. The advances in telecommunication networks and groupware systems, coupled with the growing interest in distance education, provides a synergy which has led to an increased interest in using the Web as a formal educational delivery mechanism.

The University of Pittsburgh School of Information Sciences has been actively involved in research to develop a computer-supported, collaborative learning environment [Mahling, et al. 1995]. While our initial system was a UNIX/X-Windows-based learning environment, our more recent efforts are aimed at reaching a wider audience and supporting synchronous as well as asynchronous learning via the WWW.

This paper presents our findings with respect to interface design strategies for computer-supported, collaborative, problem-based learning environments. The remainder of this paper is divided into three main parts. In section two, we introduce two educational scenarios for distance as well as collaborative learning within which our systems have been used and evaluated. In section three, we provide a chronological summary which highlights our research in developing collaborative learning environments to date. Section four discusses the collaborative learning interface requirements that we have discovered.

2.0 Computer Support for Collaborative Learning

Developing effective instructional software for Computer-Supported Collaborative Learning (CSCL) demands that it be flexible enough to accommodate various patterns of use [Koschmann 1996]. In this section, we briefly review two of those CSCL scenarios, collaborative learning and asynchronous/synchronous problem-based learning, within which our collaborative learning environment has been tested.

2.1 Collaborative Learning

The objective of collaborative learning is to encourage a group of students to work together to solve a problem. Collaborative learning strives to foster teamwork, individual accountability, prompt feedback, high self-expectations, and a respect for diversity among group members. Several studies have shown collaborative learning to be an effective model for education [McKeackie 1980; Kulik & Kulik 1979; Smith 1986]. Shared editing, synchronous and asynchronous work on a case, or navigating an information space together are some examples of opportunities where advanced computing technology can add to this pedagogical approach.

2.2 Synchronous/Asynchronous Problem-Based Learning

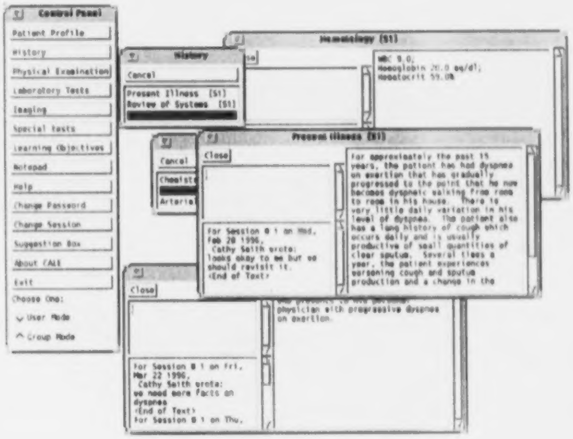
In recent years, problem-based learning (PBL) has received increased attention as a tool in medical curriculums and as the basis for designing new, innovative curricula in other fields as well. Medical schools have looked to PBL as a means to teach problem solving skills, to help students develop independent learning skills, and to create a bridge from lecture-based to more collaborative-based courses [Barrows 1994].

PBL helps students improve their reasoning skills by encouraging them to consolidate isolated facts into connected, conceptual clusters. PBL has been chiefly supported by conventional documents and "paper patient simulations", though an increasing number of computer-supported environments are emerging [Grisson & Koschmann 1995; Mahling, et al. 1995; Hmelo, et al. 1995]. We believe that electronic information technology developed for the Web can truly unlock the potential of PBL for many learners in a variety of academic domains. Multimedia enables case materials to be represented very realistically. In addition, data systems minimize the bookkeeping chores found in PBL course administration. Also, the documentation created during the group's approach to the problem can be automatically recorded. Advances in groupware research can be applied to provide computer support for cooperative, problem-based, distance learning.

3.0 Mapping Stand-alone Applications to the Web

Our research in computer support for collaborative learning began as a collaboration with the University of Pittsburgh School of Medicine. The collaboration was centered around how computers might help support the School of Medicine's efforts in implementing a problem-based learning curriculum. Finding a more efficient way to deliver PBL cases to groups of students, as well as providing tools that support and facilitate collaboration among small groups of students, were among the chief concerns.

3.1 CALE I: PBL for groups under UNIX



WBC 9.0;
Hemoglobin 20.0 mg/dl;
Hematocrit 59.0%

For Session # 1 on Wed, Feb 28 1996,
Cathy Smith wrote:
looks okay to me but we should revisit it.
<End of Text>

For approximately the past 15 years, the patient has had dyspnea on exertion that has gradually progressed to the point that he now becomes dyspneic walking from room to room in his house. There is very little daily variation in his level of dyspnea. The patient also has a long history of cough which occurs daily and is usually productive of small quantities of clear sputum. Several times a year, the patient experiences worsening cough and sputum production and a change in the

physician with progressive dyspnea on exertion.

For Session # 1 on Fri, Mar 22 1996,
Cathy Smith wrote:
we need more facts on dyspnea
<End of Text>
For Session # 1 on Thu,

Figure 1 CALE I Interface

BEST COPY AVAILABLE

CALE I (Fig. 1) is an X-Windows application and includes functionality to support synchronous as well as asynchronous collaborative problem-based learning. The system is a comprehensive, collaborative learning environment where students explore PBL cases on-line, take notes using a shared information space, and associate comments with case materials for future reference and learning by the group. CALE I was introduced as part of the University of Pittsburgh's *Medial Decision Making* course.

3.2 CALE II: Porting an X-Windows Application to the WWW

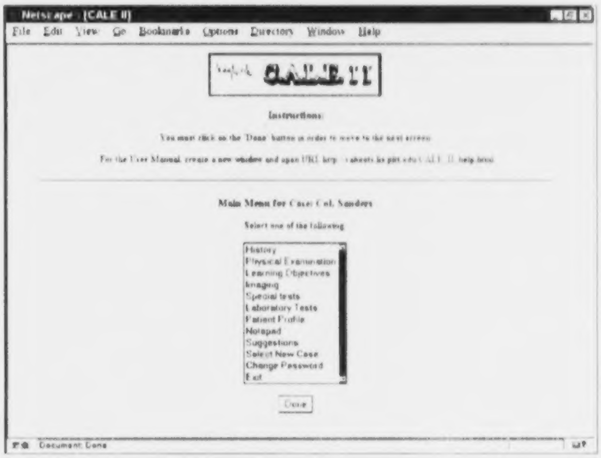

Netscape - [CALE II]
File
Edit
View
Go
Bookmarks
Options
Directory
Window
Help
CALE II
Instructions:
Main Menu for Case: Col. Sanders
Select one of the following
History
Physical Examination
Learning Objectives
Imaging
Special tests
Laboratory Tests
Patient Profile
Notepad
Suggestions
Select New Case
Change Password
Exit
Done
Document: Done

Figure 2 CALE II Interface

The ability to reach a wider audience via the WWW resulted in CALE II, a web-based version of our collaborative learning environment (Fig. 2). CALE II was used by the University of Pittsburgh Pathology Department as part of their *Integrated Life Science in Pathology* course. The CALE II interface was restricted to a single window due to the limitations in the HTML standard at that time. The single window interface strategy placed a considerable cognitive load on students as they navigated and worked through PBL cases. The insight gained from student evaluations of the CALE II interface, coupled with advances in tools for Web application development, led us to develop our current web-based collaborative learning interface (CoMMIT) (Fig. 3).

3.3 CoMMIT: A WWW PBL Interface Based on Frames

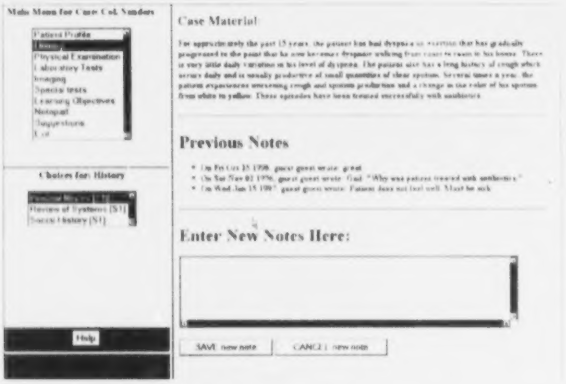

Main Menu for Case Col. Sanders
Patient Profile
History
Physical Examination
Laboratory Tests
Imaging
Special tests
Learning Objectives
Notepad
Suggestions
Exit
Choices for: History
Review of Systems (S1)
Social History (S1)
Help
Case Material
Previous Notes
Enter New Notes Here:
SAVE new note
CANCEL new note

Figure 3 CoMMIT Interface

BEST COPY AVAILABLE

CoMMIT is a frames-based web application and has been used in the University of Pittsburgh's Department of Information Science and Telecommunications as part of an undergraduate course in Human-Computer Interaction. The remainder of this paper presents our findings resulting from empirical testing and evaluation of these three interfaces.

4.0 PBL Interface Guidelines

The look-and-feel of our collaborative learning environment has changed dramatically since its inception over four years ago. The evolution of these interfaces has revealed a number of interface design issues regarding appropriate windowing strategies, menu presentation strategies and structures, and annotation mechanisms that are conducive to computer support for collaborative, problem-based learning.

4.1 Effects of Window Strategies on Case Navigation

The memorization of isolated facts proves to be ineffective for complex problem-solving tasks [Spiro, et. al. 1987]. PBL curriculums aim to overcome this on a case-level by requiring students to integrate information from several case documents to support their hypotheses or confirm their conclusions. System designers often employ a multiple-window interface strategy in situations where several information sources must be consulted simultaneously; however, the effects of single-vs.-multiple windows in computer-supported learning environments is still debated [Bly & Roesenberg 1986; Benshoof & Simon, 1993].

Each of our three prototypes employed a different windowing strategy to determine which is most effective. The CALE I system (Fig. 1) used an overlapping window strategy. Students were able to keep as many windows open as they liked; however, student response confirmed that this strategy often leads to feelings of being overwhelmed with "window-housekeeping chores" and not being able to spend enough time on the task. This finding is consistent with the findings in user-interface design research which points at the importance of letting the users focus on the domain tasks with minimal cognitive effort used for interface navigation [Card, Moran, & Newell 1983].

The web-based, CALE II interface (Fig. 2) employed a single window strategy (primarily because web development was not conducive to multi-window strategies at that time). A linear sequence of full-screen menu choices were presented to the students until the desired case material was eventually presented. Students again reported feelings of being "lost" and complained that they could not form an appropriate mental model of the case space or where they were within the case. The single-window model was clearly not appropriate. It is interesting to note that neither the total flexibility of multiple overlapping windows, nor the rigidity of single window task focus were appropriate for the learners.

The CoMMIT interface displays both the main and corresponding secondary menus at all times to facilitate students' navigation through a case. Student-requested case documents are presented in a separate tiled window. A group Notepad resides in an accompanying, floating window to support the need to organize group thoughts and ideas during case exploration. Overall, students have responded positively to a tiled-window strategy coupled with a floating Notepad window, yet simultaneous presentation of multiple documents is still a problem. We are currently extending the functionality of the CoMMIT interface to employ a combined tiled and overlapping windows approach to allow for the viewing of multiple case documents simultaneously.

4.2 Menu Structure and Presentation

PBL presents a challenge for the system designer to determine an effective way for structuring and providing access to case documents such that the system is conducive to case exploration. Students follow an iterative cycle of requesting information, analyzing and integrating this information with what is already known, and determining whether the case can be solved or if the cycle should be repeated. Supporting this high level of information requests requires the thoughtful choice from a host of menu structuring strategies. A comprehensive taxonomy of menu strategies has been suggested by Schneiderman [Schneiderman 1992].

When users have a large number of selections from which to choose, menus organized by categories are an effective strategy [Norman 1991]. Students using our system find a two-tiered menu structure with domain-specific categories on the Main Menu and corresponding case materials on a corresponding second-level menu. This strategy allows PBL case authors to help shape the students mental model of the domain and use menu terminology that is familiar to the students. Students expressed that this two-tiered menu structure works well, but only if the menus are visible at all times.

The single interface of CALE II required us to modify the presentation of our two-tiered menu. In CALE II, students would first be presented with a list of Main Menu options. After selecting an item from the Main Menu, the system next presented the second-level menu with options corresponding to the Main Menu choice. Choosing one of the secondary menu options resulted in a case material being presented. After students were finished viewing a case material, the system would return them to the Main Menu and the cycle would begin again. This linear presentation of two-tiered menus resulted in a substantial cognitive overload for students. Students expressed feelings of "getting lost" in the menu structures and felt that it was difficult to form an appropriate mental model of the case document space.

We have found that the availability of the menu at all times is critical for collaborative, PBL environments. In the CoMMIT interface, students can see the Main and Second-level Menus at all times - each menu resides in a separate tiled window. Choosing an item from the Main Menu updates the Second-Level menu. Selecting an item from the Second-Level Menu presents the selected document in the case material window. Using this menu model allows students to see how they got to a particular case material and reminds them of the document categories from which they can choose.

An interesting student behavior that we observed is that students rely on the menu not only as case material selection mechanism, but also as a mechanism for coordinating group activities. Student evaluations of all three interfaces suggest that the menu structure should include status information such as which options were attempted by the group, whether or not the request was successful or not, and if not, how many times had the option been tried. We look forward to incorporating these suggestions into our next version of CoMMIT and assessing its utility in facilitating group problem-solving.

4.3 Context-sensitive classification of Case Annotations

Students in paper-based PBL often use a physical blackboard divided into four columns: Facts, Hypotheses, Learning Issues (to do's) and Actions to help organize the group's thoughts and ideas during case exploration [Meyers, et al.1990]. In the paper-based PBL environment, one student in the group acts as a scribe to record and update the information in these four categories on the blackboard as the group proceeds through the case. To support this requirement in our system, we developed a shared information space called the NotePad that follows the blackboard metaphor. During case exploration, students switch to the Notepad Window to record information in any of the four categories as the need arises. The system records the name, time, and date of student annotations and orders those annotations from most-to-least recent.

We found that the students perception of the blackboard metaphor changed when the blackboard was implemented electronically. Students suggested that while the blackboard metaphor provided some computer-support for the group's information needs, they preferred to enter this information directly with the case material rather than using a separate Notepad window. We found that students often used our Margin Note feature (originally intended for making only general comments "in the margins" of displayed case materials) in lieu of the Notepad when entering information about facts/hypotheses, etc. This practice was consistent across all groups in all domains that used our system.

An analysis of this phenomenon led us to conclude that supporting the Margin Note approach to group case annotation has several merits:

students remain focused on the task of annotating rather than concerning themselves with window management tasks of switching back and forth between the case material and Notepad windows.

annotations used with the Margin Note feature provided a richer context within which to understand student annotations thus students were more likely to annotate for both themselves and for the benefit of the group.

because the annotations were more contextually dependent, facilitators could more accurately assess the breadth and depth of the students' knowledge and reasoning which is a fundamental principle of PBL [Koschmann 1996].

annotations made with the case material can be classified by the students at entry and automatically indexed in the NotePad such that group activities can be viewed at a glance. In this way, the NotePad can serve as a point of departure for future collaborative sessions on the case.

5.0 Summary of Lessons Learned

Our initial efforts at supporting collaborative learning in a problem-based learning environment were concerned with providing a shell within which PBL cases could be delivered to groups of students. Although our initial systems did provide computer support for collaborative problem-based learning, our experiences in implementing three different interfaces helped us uncover more subtle interface requirements for this type of learning environment. Specifically:

students prefer a semi-structured window management strategy over a totally unstructured or totally rigid window management scheme,

a two-tiered, hierarchical menu structure is effective for students to form and maintain a mental model of the case document space but only if those menus are displayed together and at all times,

the blackboard metaphor is only partially effective in our computer-supported PBL environment. Students prefer to organize hypotheses, facts, and action items at the point of entry (with the case materials themselves) rather than using a blackboard metaphor,

a sorted, centralized compilation of student case material annotations (done automatically by the system) provides a high-level perspective of group activities. These centralized compilations can permit both students and facilitators to more accurately audit the group's knowledge and problem-solving processes.

References

[Barrows 1994] Barrows, H. S. (1994). *Practice-based learning: problem-based learning applied to Medical Education.* Springfield, IL: Southern Illinois University School of Medicine.

[Benshoof and Simon 1993] Behnsoof, L. A., & Simon, H. (Autumn 1993). The Effects of Single and Multiple Window Presentation on Achievement During Computer-Based Instruction. *Journal of Computer-Based Instruction,* 20 (4), 113-117.

[Bly and Rosenberg 1986] Bly, S. A., & Rosenberg, J.K. (1986). A Comparison of Tiled and Overlapping Windows. *CHI '86 Human Factors in Computing Systems Conference, 1986.* Association for Computing Machinery's Special Interest Group on Computer and Human Interaction (ACM/SIGCHI) in cooperation with the Human Factors Society and ACM/SIGGRAPH, Boston, MA..

[Card, Moran, & Newell 1983] Card, S. K., Moran, T.P., & Newell, A. (1983). *The Psychology of Human-Computer Interaction.* Hillsdale, NJ: Lawrence Erlbaum Associates.

[Grisson and Koschmann 1995] Grisson, S., Koschmann, T.D. (1995). Hypermedia without Programming: Automatic Generation of Presentation Documents for Case-Based Instruction. *National Educational Computing Conference 1995,* Eugene, OR.

[Hmelo, et al. 1995] Hmelo, C., Vanegas, J., Realff, M., Bras, B., Mulholland, J., Shikano, T., & Guzdial, M. (1995). Technology Support for Collaborative Learning in a Problem-Based Curriculum for Sustainable Technology. *CSCL '95* Bloomington, IN. 169-172.

[Koschmann 1996] Koschmann T. (Ed.). (1996). *CSCL: Theory and Practice of an Emerging Paradigm*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

[Kulik and Kulik 1979] Kulik, J. A., and Kulik, C.L.C. (1979). *College Teaching.* In P. L. Peterson and H. J. Walberg (Eds.) , Research on Teaching: Concepts, Findings, and Implications (pp. 70-93). Berkeley, CA: McCutchan Publishing Corporation.

[Mahling, et al. 1995] Mahling, D., Sorrows, B., and Skogseid, I. (1995). A Collaborative Environment for Semi-Structured Medical Problem-Based Learning. *Computer-Supported Collaborative Learning, 1995,* Bloomington, IN. 218-225.

[McKeackie 1980] McKeackie, W. (Ed.). (1980). *Learning, Cognition, and College Teaching.* San Francisco, CA: Josseey-Bass.

[Meyers, et al. 1990] Meyers, A., Barrows, H., Koschmann, T., and Feltovich, P. (1990). The Facilitation of Problem-Based Learning in Medical Education Through a Computer-Mediated Tutorial Laboratory. *Symposium on Computer Applications in Medical Care, 1990.*

[Norman 1991] Norman, K. (1991). *The Psychology of Menu Selection: Designing Cognitive Control at the Human/Computer Interface*. Norwood, NJ: Ablex Publishing Corporation.

[Schneiderman 1992] Schneiderman, B. (1992). *Designing the User Interface* (2nd ed.). Reading, MA: Wesley Publishing Company.

[Smith 1986] Smith, D. G. (1986). *Cooperative Learning Groups.* St Paul, MN: Northwest Area Foundation.

[Spiro, et al. 1987] Spiro, R. J., Vispoel, W., Schmitz, J., Smarapungavan, A., and Boerger, A. (1987). *Knowledge acquisition for application: cognitive flexibility and transfer in complex domains.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Experiences in Virtual Teaching

Karen A. Lemone
Associate Professor
Computer Science Department
Worcester Polytechnic Institute
Worcester, MA 01609, USA
(508) 831-5529
kal@cs.wpi.edu

Abstract: Virtual teaching via the Web is becoming commonplace. Tools to better enable this activity are beginning to appear. However, little formal assessment has been done to determine effectiveness of such tools nor the effectiveness of such distance learning. In this paper, we describe experiences teaching online Web courses and a set of formal assessment procedures for evaluating such courses. The courses, the tools and the assessment procedures have evolved over multiple teachings of the same two courses over the last three years both in the US and in Europe. One course is a Web *Publishing* course for non-computer science majors. The other is a Web *Programming* course for computer science majors. Statistics for both graduate students and undergraduates are included.

1. Introduction

Teaching courses via Web materials has new teaching issues plus old issues in a new setting: Just as in traditional courses, TA's and other assistants are needed. Traditional tasks (officer hours) as well as non-traditional tasks (staffing Chat Rooms) are needed. They are needed for maintenance of class pages, answering student questions - asynchronously via email and synchronously by holding "office hours" in Chat Rooms.

While routine homeworks can be graded, recorded and responded to automatically, good software tools to enable this are just being developed. We have just developed and tested such tools. In the versions of the courses assessed here, all homeworks were graded by hand electronically and results emailed to the students. As will be described, this does not work well.

When instructors teach a course for the second (third, fourth, ...) time, they reorganize existing material to make it appropriate for the current class. In traditional mode, this may include adding and deleting material, creating new projects, quizzes and assignments, refocusing for a different audience, etc. We have developed software to facilitate these tasks, but have yet to test them. Thus, all changes to these courses from previous versions was done by hand, checking and editing the course pages.

The Web provides poor facilities for searching and navigating. Supplemental tools were developed and used in summer 1997 for the first time.

We group our tools into a system called *ReCourse* [Lemone, 1996]. It also has been evolving over the last three years. It is a Web Retargetable Course Generation System whose purpose is to facilitate both distance and on-campus learning via the World Wide Web. By "retargetable", we mean the process of changing the Web course to "target" it for a different term or audience.

ReCourse's features include:

- Ability to retarget a Web course for different levels of students. A user-friendly editor allows instructors to add appropriate tags to HTML documents. Students then see only the parts of the pages appropriate for their level.

- Multiuser chat rooms to facilitate synchronous student, instructor, and TA communication.
- A secure grading system allowing instructors to record grades and students to view their own grades.
- Bookkeeping Tools such as a Hypertext Link Check to ensure that all internal and external hypertext references are valid, Search facilities, and Content Update tools to allow global updating of course pages (e.g., changing the term and date headers, course icons etc.)
- A Map Generator to create a semi-static site map of the pages to allow students a birds-eye view of where they are in the course pages. This tool is run periodically by either the instructor or TA's when changes have been made in the organization of the course pages.
- A Quiz Feedback system.
- A course bulletin board (news group).

The system can be entered as an administrator who installs the tools, as an instructor who sets up things such as the grading pages, generates the site map, checks for dead links etc., or as a student who can access the news group, the site map, his/her grades etc.

This paper reports on the results of teaching using these tools, rather than on the tools themselves. More information on the tools can be found at http:/.www.webrecourse.com.

2. Instructional Model

People have been teaching courses via the Web for a number of years now. Sometimes the Web is used as a supplement to the class. Sometimes it is where the class takes place. We have experimented with a number of models and instructional designs and have learned and are still learning about the impact on student learning and faculty productivity of these models. In this paper, we describe results of teaching two summer courses almost entirely online. There was one meeting at the beginning where students met each other and the instructor, and the course format was discussed. At a final meeting at the end of the course, students presented the projects they had created during the course.

A pretest was administered at the first class and a posttest with the same questions was administered at the final meeting. We describe these assessments and their results.

2.1 Instructional Design

ReCourse is a Web-based system used in conjunction with Web course pages. It presumes course pages exist in a directory, and that there is a "root node" (home page); other pages are connected as links in the typical web-like architecture. Future enhancements can facilitate this creation. A typical course would have a number of modules representing the major topics in the course. Links also exist to the course information - email and phones of the instructor, TA and graders, Syllabus, Class list - with references to their home pages (if any) and their email addresses - Project decription (if any), and grading.

The two courses, *Network Publishing* (http://cs.wpi.edu/~kal/netpub) , a Web Publishing course for non-computer science majors, and *Electronic Documents* (http://cs.wpi.edu/~kal/elecdoc), a Web Programming course for computer science majors were similar in format: a number of modules of information for which they sent in weekly homework, weekly labs which taught the publishing (page creation) and programming (Client and Server languages) aspects of the course, and a significant project which could be done singly or in groups.

2.2 Educational Technology

Although the Web courses may be used within the classroom structure, they were developed for a distance learning model. Having taught this way for three summers, we have developed and incorporated techniques to facilitate distance learning: multiple (Web) references and weekly homeworks for reinforcement of the material, personalized responses when homework is submitted, and "presence" (asychronously via email, synchronously via Chat Rooms). In addition, the tools include automatic feedback on homework and birds-eye views of pages so that students can see where they are in the material and find other information more quickly.

2.3. Comparison with Other Instructional Models

Non Web-based distance learning models have relied on videotapes and broadcasts. While some Web courses have been taught synchronously via White Boards, etc., the technology just isn't sufficient yet. Our model is primarily asynchronous, allowing both the instructor and students to work at their own place, rate, and time. Our assessments included questions evaluating these features.

Most Web-based courses are created and maintained by the instructors, perhaps with TA help. Few systems exist to aid the teaching of Web courses. WebCT [Goldberg 96, http://homebrew.cs.ubc.ca:8080/] comes the closest to ReCourse, but it lacks the "retargeting" facilities: when a course is retaught, it needs to be changed, updated, etc. Web courses take a phenomenal amount of time to develop, update and maintain. Tools to reuse material are needed. We know of no other system that addresses this retargeting issue.

It was our hope that productivity would improve for the instructor and students due to:

- TA help in chat rooms, a bulletin board and email. We spent hours each week, responding to email in the past. Sometimes, we could not respond in a timely manner. Support personnel are needed for distance learning in many of the same ways that they are needed for traditional classes. In fact, students may need *more* online personal contact from course personnel to overcome the lack of personal presence. The bulletin board was not ready for the summer, and perhaps because of this, the email quantity was again a major problem for the instructor and staff.

- Automatic grading of weekly homeworks. We use routine assignments to encourage reading and assimilating of the course material. We grade them ourselves and send students feedback and their scores via email. Again, this takes a few hours/week. The automatic test system will ease this. We did not have this fully tested and integrated for security this summer, but it will be used this Fall. The conclusions will discuss the very real need for such a system as well as a potential drawback.

- The Bookkeeping Tools allowed the instructor to quickly find dead links, and to generate a site map; students were able to use this site map to “see” where they were in relation to the rest of the pages. The search tool (suggested by a previous class) was extensively used.

- The retargeting tools will enable the instructor to create the next version of the class in far less time than we presently spend. They were not used for the summer versions assessed here.

- Instantaneous feedback to students on their homework. For this version, just a personalized acknowledgment page was sent; the next version of the course will send back a graded page with correct answers and a paragraph of explanation for each question. Issues of security (the answers were accessible via a Java program) prevented their use this summer.

- Automatic and secure access to student grades (for students and the instructor.) Again, this was not fully secure for the summer, and students expressed a strong desire for it.

3. Assessment Plan

We were funded by the Davis Educational Foundation to develop and perform statistically significant assessments on these classes.

3.1 Procedures and Instruments to Measure Effectiveness

We have been using student questionnaires for the last 3 years. There is a preliminary questionnaire, and a post questionnaire for each course. One term, students filled out weekly assessments. Interestingly, students have always filled out these electronic Web forms even when they ran a week or two behind. We've never gotten anywhere near this response with paper questionnaires!

However, we decided more formal assessment procedures were needed.

3.2 Description of Control Groups and Comparison Tools

We assessed the effectiveness of the Web courses and the ReCourse software in the summer versions of two classes: Electronic Documents and Network Publishing. The Network Publishing group are less technical, more writing and publishing-oriented (in theory). The Electronic Documents group are Computer Science or Computer Engineering majors (or those with strong computer backgrounds.) We compared these groups, not with each other, but with information gathered via a pretest and a posttest. We gathered and compared issues such as (1) background, (2) behavior, (3) attitude, (4) satisfaction, and (5) knowledge and skills gained.

3.3 Pre/Post Analysis

For the preliminary questionnaire, we asked questions about their background and interests, e.g., questions concerning Web experience. For behavior, we asked questions such as the number of hours per week they planned to spend. For attitude, we asked questions such as whether they (would/would have) prefer/red the course to be taught in the traditional manner (as opposed to online). For satisfaction, we asked questions such as helpfulness of the instructor and whether they think/thought the course to be useful.

Finally, both the pretest and the posttest included 100 objective (mostly multiple choice) questions relating to the material. Because of the large number of questions, it was hoped they would not remember a significant number of questions when studying for the posttest.

We also used the WPI standard course evaluation form (The first 14 questions indicate an overall measure of satisfaction, and the very last question indicates self-perception of learning.) These results are not yet available.

4. Outcomes

We summarize the results of the various categories.

4.1 Measurable Outcomes

Background: Not surprisingly, the non-computer science majors showed less preliminary knowledge of Web related information: few had created Web pages although most had used the Web. About 3/4 of the computer science majors (Electronic Documents course) had a Web page, and about ¼ indicated some knowledge of client and server programming languages (primarily Perl, JavaScript, and Java)

Behavior: On the pretest, most students indicated they planned to spend 15-20 hours/week with a few planning fewer hours. On the posttest, the majority indicated that they spent in excess of 20 hours a week with a few spending less.

Similarly, most students expected to spend 3 days/week before the course, but indicated having spent 4 or more on the posttest.

Students were split on the pretest as to whether they planned to print out the course pages or not; most indicated on the posttest that they did print out at least some of the pages.

On the pretest, students were split between spending 5-10 hours and 10 or more hours "surfing" the web. The totals were actually down in the posttest.

Most students didn't know whether they would use the chat room or not before the course. Most of the more technical students in the Web Programming course said they did not use it, while many of the less technical Network Publishing students used it more - they also came to the "in person" office hours. Both the TA's and the instructor used the chat room, and they all indicated they thought it was an effective way to deal with students.

Attitude: Most students "liked the idea" of taking a course online as opposed to the traditional in class model as indicated on the pretest with a few circling "not sure." On the posttest, *everyone* indicated they like it with 1 student indicating he/she "wasn't sure" he/she would take such courses in the future. Everyone else wanted to take more such courses. Students indicated on both the pretest *and* the posttest that they did not believe the course could be done with no meetings at all.

Satisfaction: Most, but not all students indicated that the course objectives were clear both before and after the course. Almost everyone felt the course was well organized. Most, although not all, students expected and found the material challenging and interesting. Not everyone felt the instructor was helpful, while most expected her to be so. Everyone expected to be able to apply the materials and skills learned to their professional lives. Most, although not all, felt the homeworks and the assessment (posttest) measured their knowledge of the material. Only one student felt he hadn't learned a lot in the course.

Course Material: No one knew many of the answers for the pretest. Posttests were, of course, much better although it will be interesting to compare these results with those of the next course (none of the tests are allowed to circulate.)

5. Conclusions

Class satisfaction has been high in the past, and continued to be so. Students seem to like taking a course (mostly) on their own in the summer. Whether this model would be successful during the year or for many of their courses remains speculation. Although not as objective as times and correct answers to a question, satisfaction can still be measured, at least qualitatively, and reported on. Comparison of the student's desired outcome ("What do you hope to learn in this course?") described on the pretest with the actual outcome ("Did you learn (less than/more than/ etc.) what you hoped to learn") on the posttest, is an important measurable. (We email back right away when a desired goal is unrealistic for the course.)

Nevertheless, the formal assessment procedures indicated possible areas of improvement. Given that the instructor was spending many, many hours/week on the course, it was disheartening to find out that some students felt they were not able to communicate well. A course bulletin board, better grading software and a better delegation of tasks among ta's and instructors may improve this.

In the past, both students and instructors have spent more time on these courses than on traditional in class courses. The ReCourse tools have improved things somewhat, but the amount of time still appears excessive. Further tool development (the bulletin board), and reorganization of the modules may improve this for students. For example, both courses included a module on the theory of hypertext which should likely be spread between two modules. Having a staff did improve instructor efficiency, but more improvements are needed. (The two courses still took in excess of 20 hours/week!)

When the "retargeting facilities are fully integrated, time spent prior to the course should decrease.

Most of all, the automatic grading and recording of homework will significantly decrease instructor time as well as providing less "human error" in grading. Whether this will result in further student alienation will have to be assessed.

These are important outcomes. If online Web courses are to be taught in the future, appropriate tools need to be made available and assessment should measure whether students are learning and satisfied with the way they are learning.

6.0 References

[Agonino and Wood 1994] Agogino, A.M. and W. H. Wood III (1994). The Synthesis Coalition: Information Technologies Enabling a Paradigm Shift in Engineering Education, keynote talk, in Hyper-Media in Vaasa'94. *Computers and Hypermedia in Engineering Education* (Ed., M. Linna and P. Ruotsala,), Vaasa, Finland, 3-10.

[ARIADNE 95] Alliance of Remote Instructional Authoring & Distribution Networks for Europe, Project Overview, EU Framework 4 Project, *Telematics Applications. Education and Training*, 1995.

[Lemone 96] Lemone, K. (1996) Retargetable Course Generation, A Methodology for Reusability, *Workshop on Architectures for Intelligent Tutoring Systems*, ITS '96, Montreal, Canada, June 1996.

[Lemone 97] Lemone, K. (1997) Assessment of Tools for Virtual Teaching, *Enable '97*, Espoo, Finland, June 1997.

[Lemone 97] Lemone, K. (1997) WebRecourse, http:www.webrecourse.com.

[Goldberg 96] Goldberg, M. (1996) WebCT, http://www.webct.com.

[Goldberg and Salari 97] Goldberg, M. W. & Salari, S (1997). An Update on WebCT (World-Wide-Web Course Tools) - a Tool for the Creation of Sophisticated Web-Based Learning Environments. *NAUWeb '97 - Current Practices in Web-Based Course Development*, June 12 - 15, 1997, Flagstaff, Arizona.

[Watabe, Hämäläinen, and Whinston 95] Watabe, K., Hämäläinen, M., & Whinston, A. An Internet Based Collaborative Distance Learning System: CODILESS, Computers in Education, Vol.24, No.3, 1995,141-155.

6.1 Acknowledgments

I am indebted to Sean White who implemented much of Web ReCourse and to Sean White and Yong Li for their superior assistance in the summer 1997 versions of Network Publishing and Electronic Documents.

A Taxonomy of Internet Communication Tools

Byron Long and Ronald Baecker
Dept. of Computer Science and Knowledge Media Design Institute
University of Toronto
10 King's College Road
Toronto ON Canada M5S 1A4
E-mail: byron@dgp.toronto.edu, rmb@dgp.toronto.edu

Abstract: When looking at the social phenomenon that are arise from the use of Internet communications tools, one must consider the properties of the tool that influence human-to-human interaction. This paper presents a number of such properties and discusses their importance. In addition, existing Internet communications tools are described both in general and with respect to these properties.

Introduction

The term 'virtual community' has been used to describe all manner of computer supported communication. In some cases the sum total of all such communications is termed *the* virtual community, but in most cases the term is limited to communication that makes use of a single network resource. But the ability to communicate alone does not ensure that a community will form. Indeed most attempts to define exactly what comprises a virtual community require an in depth look at what is required for a connection to become a community. Often such definitions are presented as a collection of anecdotes that attest to the social diversity necessary for 'community' [see Morningstar 91 and Rheingold 93].

Virtual communities that are in existence today are supported by a wide variety of communications tools. The various properties of these tools exert a strong influence on the character and structure of the communities they support. An examination of these tools can be cast in terms of the properties that most shape communities built with them. This paper both distinguishes properties which have a significant bearing on social interaction, and describes the various categories of tools for communication on the Internet.

Properties

Conversational Synchronization

An important distinction between these tools is the synchronization between the composition of a message and its receipt. The Internet was designed to support *store-and-forward*, or asynchronous methods of communication. In this type of communication, any one message is received at some interval after it has been composed, usually when it is explicitly requested. In most such systems, particularly email and news, this results in the receiver of a message perceiving that the sender is more intelligent or eloquent that would otherwise be the case. This perception arises out of the increased amount of time that can spent composing an effective message.

Real-time, or synchronous, communication, on the other hand, does not allow for extended delays in message composition. Applications such as Internet Relay Chat, video conferencing and Internet telephony require that participants respond in turn to their conversational partners' utterances. This leads to an experience more similar to face-to-face conversation than the store-and-forward exchange of letters.

Some real-time methods use text as the medium of communication, which allows one to trace the history of a conversation with some accuracy, while others use audio and video, where the specifics of conversation are ephemeral, and must be recalled by participants.

Conversational Style

Another property of computer-mediated communication is the conversational style that each method supports. Email and Talk support a *person-to-person* style of conversation, where both conversants are equal partners in the exchange.

On the opposite end of the scale are the web, Internet radio, and FTP, which are *broadcast* media. The composer of the message sends it out to many people, most of whom are unable to respond in the same medium, and those that can are generally unable to directly target the original sender.

As a median between these there are *forum* style methods of communication. Examples of such a style are newsgroups, electronic mail lists and a large number of real-time conversation ('chat') systems. Forums allow for conversation among groups of people, with each person being able to respond to each other participant.

Communications Media

Another distinguishing feature of communications applications is the conversational media they support. Most systems support text, the original media of Internet communication, though some (Internet radio and Internet phone) support only audio. A growing proportion of Internet traffic includes static graphic images, as supported by the web, and chat systems implemented on the web (WOOs and web chats) generally also support limited graphics - generally pictures of the conversants. A limited number of applications support a representation of each participant in the conversation. These 'avatars' [Morningstar & Farmer 90] allow for the positioning of a participant within the setting of the conversation (*Virtual Places*) and can also represent the person's facing (most virtual environments, including *WorldsChat*).

Some media, audio and video in particular, are highly ephemeral. Communication requires active attention or conversational flow is lost. Most other media, however, leave a short term trace of recent utterances and therefore can support a more detached conversational participation.

Initiation Method

The different tools support a number of ways in which conversational partners locate one-another. For some tools, like newsgroups, which propagate messages through *replication*, no effort is required on the part of the user; the messages are simply available, and they merely need to add their own contribution. Email on the other hand, requires that a message writer know the user name and machine name of their reader's Email account, their Email *address*. Other types of addresses, such as *ICQ* numbers, also exist, serving as an indirect indicator of user name and machine.

Many real-time tools (Internet radio, video conferencing, Internet phone, etc.) require that the connection be made to the *machine* that the other conversant is using, through the machine's address. Others require connection by all participants to a single *server*, also based on address. In such a case, all communication is routed through the server machine.

For some tools (*Virtual Places*, Web News) the space of conversation is defined by a particular World Wide Web page, located either by browsing through the web or by using a specific *URL*.

Locating other conversant through a server, a URL or replication doesn't require that a participant previously know the others they communicate with. Mutual knowledge of the location of a communication resource is all that is required to be a member of the community.

Audience Membership

Some applications require that participants be members of a certain system, rather than being part of the global membership of Internet users. On such *constrained* membership systems, one can only communicate with other members. BBS's, in fact, allow for the use of Email, newsgroups, and chat systems resembling IRC among a constrained membership, rather than the *global* membership supported by the individual tools.

Having a constrained membership leads to more personal accountability. Disruptive acts are more easily tied to an individual, and such acts can put an individual's group membership in jeopardy.

Dialog History

For many of these tools no history of the conversation which has taken place up to the current point in time is available. Without a history of communication, a new participant in the conversation is unable to acquaint themselves with the conversational style of the other participants and with the recent course of discussion. While for some tools this is not a problem, either because they are person-to-person, based on real-world conversational protocols, or they have no salient course of discussion, for others it can be problematic.

When there is not dialog history, the arrival of a new participant is often marked by a period of introduction where the newcomer attempts to get up to speed. This requires a fair deal of social initiative however, and many newcomers must 'lurk' for a time before feeling sufficiently grounded to participate.

Tool	Synchronization	Style	Audience Membership	Communications Media	Dialog History	Initiation Method
Email	Store-and-forward	Person-to-Person	Global	Text	No	Address
Newsgroups	Store-and-forward	Forum	Global	Text	Yes	Replication
FTP	Store-and-forward	Forum	Global	Text	Yes	Server
Web News	Store-and-forward	Forum	Global	Text	Yes	URL
Email Lists	Store-and-forward	Forum	Constrained	Text	No	Address
Collaborative Hypertext	Store-and-forward	Forum	Constrained	Text, Graphics	Yes	Replication
World Wide Web	Store-and-forward	Broadcast	Global	Text, Audio, Graphics, Video	-	URL
Internet Radio	Store-and-forward	Broadcast	Global	Audio	-	Server
Shared Whiteboard	Real-time	Person-to-Person	Global	Text, Graphics	Yes	Machine
PowWow	Real-time	Forum	Global	Text, Audio	No	Machine
Virtual Places	Real-time	Forum	Global	Text, Audio, Graphics, Avatar	No	URL
Virtual Environments	Real-time	Forum	Global	Text, Audio, Graphics, Avatar	No	Server
Talk	Real time	Person-to-Person	Global	Text	No	Machine
Internet Phone	Real time	Person-to-Person	Global	Audio	No	Machine
IRC	Real time	Forum	Global	Text	No	Server
Web Chat	Real time	Forum	Global	Text, Graphics	Yes	URL
Video Conferencing	Real time	Forum	Global	Audio, Video	No	Machine
MU*	Real time	Forum	Constrained	Text	No	Server
WOO	Real time	Forum	Constrained	Text, Graphics	No	Server
Internet Pager	Both	Person-to-Person	Global	Text	No	Address
ICQ	Both	Forum	Global	Text	No	Address
Agora	Both	Forum	Global	Text	Yes	URL
Bulletin Boards	Both	Forum	Constrained	Text	•	Server

Table 1: Properties of Internet Communications Tools

Available Tools

Although new tools for Internet communication are always being released, most can be grouped into a limited number of categories. The following listing attempts to cover as many tools as possible, though the listing is probably not complete. In some cases, a single product will sport a number of separate tools, retaining each tool's strengths an weaknesses.

Each of the following groups of tools can be categorized in terms of the properties given in the previous section [Tab. 1]. Some tool groups are represented only a single product (indicated by *italics* on the table).

Email: Email allows text messages to be composed and then sent to an individual or series of individuals. Each message passes through a number of machines until it comes to rest on the machine that hosts the recipient's mail, where it remains until it is explicitly retrieved by the recipient. This is the oldest form of communication on the Internet, originally making use of simple machine-to-machine copying and explicit delivery paths.

Newsgroups: The Internet newsgroup system allows for text messages to be sent to a newsgroup, usually focused around a certain issue or topic of discussion. This allows for people to choose which type of messages they wish to read and reply to. News articles are stored in a single place on a local server, and updated through a file replication scheme where each machine copies new articles to all other connected machines. Thus articles spanning a period of time are always available for perusal. This allows for 'casual' readership of newsgroups, where someone might occasionally check a number of newsgroups for articles whose subjects look interesting. This also allows new readers to trace back through the recent history of discussion in order to get a feel for the conversational style found among regular contributors to a certain group, giving them an opportunity to integrate themselves into the conversation inconspicuously.

FTP: Although rarely thought of as a communications tool, messages stored in files and in the names of directories allow users of these file repositories to communicate in rudimentary fashion with one another. Generally used by members of an underground file repository to make requests for certain files, or to tell others about other such repositories, the messages are usually written in a shorthand jargon in order to take as little space as possible.

Email lists: Email lists, like newsgroups, are organized around a topic, but are not as widely available, nor do they support occasional readership. By leveraging of the email protocol, lists redistribute any single message among all subscribers, so each message becomes part of members' email. If one is not explicitly subscribed to a list, it is not possible to read any of the articles, though it is possible to blindly send a message to it. Thus readership is constrained to a known group of list subscribers.

Web News: Discussion groups can also be hosted within web pages. This usually involves the use of CGI (Common Gateway Interface) scripts on the server that handle the various aspects of maintaining a threaded discussion group. To the user, it appears as though the various messages are contained within a web page.

Collaborative Hypertext: A number of GroupWare systems also distribute articles, but instead of reaching an Internet-wide audience, the participants are members of an organization. Corporate memory systems such as *Lotus Notes*, and educational systems such as *CSILE* [Scardamalia & Berieter 91] are examples of such systems. In addition to text, most such systems also support graphics, and some support considerable more diverse media types, including applets. A growing number of such systems use the Internet as a means of interconnection, and theoretically, given the appropriate access, could be used by anyone on the Internet.

World Wide Web: The web is used as a broadcast medium used by people who construct web pages representing their interests or themselves and make them available for browsing by other web users. Businesses and organizations use the web to advertise their presence and provide information. These messages can make use of text, graphics, video, audio, and any of the other growing number of media of the WWW.

Internet Radio: Internet radio tools provide the means to playback a stored sound file without having to bring all of it down from the server on which it resides. This allows for a broadcast similar to AM radio, except that specific content can be heard by an individual at any time, rather than the set times enforced by a scheduled radio broadcast. It is also possible to listen to live broadcasts with these tools, if the content needs to be up-to-the-minute.

Shared Whiteboard: Internet whiteboard applications allow two people to view a shared drawing space. In addition to simple graphics, writing on the board can be used for communication, though whiteboard applications are generally combined with other Internet communications systems, particularly video conferencing applications. There are a large number of protocols and specific applications used for shared whiteboards, some of which are commercial, and many more of which are limited use academic systems.

PowWow: PowWow is another tool for communication between web users. However, a connection must be made explicitly between two or more *PowWow* users, at which point they are able to communicate using text or audio, and are able to direct one another to web pages.

Virtual Places: Virtual Places allows people to see others that are visiting the same web page as they are. Each person using *Virtual Places* is represented by a small graphic, generally a picture of a head, which has a position within a web page. By manipulating the position of the head (a sort of Avatar), a user can take advantage of 'virtual furniture' within a web-page, to put themselves into virtual vehicles in order to participate in tours, and to initiate conversations by placing themselves adjacent to others. When two avatars are beside one another, they can communicate either using text, or, if there are only two participants, using audio through an Internet phone connection.

Web site tours can be initiated by anyone; a small vehicle appears and anyone who has moved their representation onto the vehicle when the tour operator moves to another web page moves to the new page with them. Tour members can engage in conversation with one another, but cannot explore pages not visited by the tour operator.

Virtual Environments: A new class of communication tools presents the user with a virtual space in which to communicate. One such tool, *WorldsChat*, presents users with a first-person three-dimensional world through which they can navigate [Damer, et. al. 96]. As they navigate through a virtual space divided into a series of rooms, they are able to see others exploring the space, and if they get sufficiently close to, and are in the same room as, the representation of another user or users, they can converse with those people. By providing a three-dimensional representation of the environment and the users, clues such as the facing of others can indicate what they might be focusing on, which could be a message left by someone else on a wall, or another participant, for instance.

A large number of multi-player games also fall into this category. Although not all support voice communications, they all represent the player in the space defined by the game. Although the primary purpose of the space is game play, all provide means to communicate with other players. The avatars supported by multi-player games can either be two or three-dimensional, depending on the structure of the game.

Talk: Talk is a simple system where two people can see what one another are typing; basically a formalization of a number of screen mirroring techniques that allowed this type of communication to occur on early Internet systems. It is the only text-based real-time communication system that shows the typing of another as it happens; all the others send a sentence after it has been completed, allowing for editing within a single utterance.

Internet Phone: Internet Phone applications allow audio connections between two people. Audio compression techniques allow for conversations to take place with only slight delays at each end, even with low-speed Internet connections. Connecting to an individual requires their Internet address, though users can be found through other means, such as through IRC.

IRC: Internet relay chat is a system in which groups of people can communicate with each other using real-time text. IRC servers have a worldwide usership, with individuals attending to one or more of thousands of 'channels' generally based on subject of interest. Each channel can have its own culture, including known veteran members, conversational styles, and automated participants known as 'bots' [Reid 91]. This often makes it difficult for a new member to become an equal participant within a channel, a problem found with many of the tools where the history of communication is not open to examination by new users.

Web Chat: Web-based chat systems are similar to single channel Internet Chat systems, except that they occur within a web page and thus can support limited graphic communication, generally used to include pictures of the conversants. Originally, limitations in the web protocol did not allow for automatic transmission of new utterances, so explicit requests for conversation updates were required. Some browsers now support timed or server-driven updating, and the use of new interactive technologies such as *Java* and *MacroMedia Shockwave* has resulted in a more dynamic (and natural) systems. There are fair number of these newer tools, including *Gamelan Chat* and *talk.com* which are implemented as Java applets, and *Ichat*, which is implemented as a browser plug-in.

Some web chat systems, like *WebTalk* [Donath & Robertson 94] are designed to give an awareness of others in an arbitrary web page, rather than having a web page dedicated to the tool.

Video Conferencing: Video conferencing applications such as *CUSeeMe*, allow for audio and video communication across the Internet. Generally such connections are person-to-person between anyone on the Internet with appropriate hardware, though forums can be set up by using a reflector, where everyone connected can be seen by anyone else connected to the same reflector. Unfortunately if such a group gets too large, the video can become excessively slow to update, and voice communication can break up. The bandwidth and synchronization required by video is significant, and it can often be difficult to maintain an efficient person-to-person communication on the asynchronous packet-based Internet, let alone maintain multiple connections.

MU:* MU* is generic term for a series of systems which include MUDs, MUSHes, MOOs and MUSEs, among others. Each of these systems allows one to explore around an imaginary space and to communicate with other people that are encountered within the space. Most MU*s are limited to text as their only medium, though this does allow for a much simpler construction of the spaces, as they need only be described. When a number of people are in the same space, they can talk to one another and perform simple actions, and the room often becomes very similar to a channel in IRC except that rather than gathering based on subject interest, conversations arise among those in virtual proximity. This encourages exploration of the space, which might either be constructed by a select few or may be constructed by all the members of a system.

MU*s have been extensively examined as social constructs [Turkle 95]. A large range of social phenomena have been studied within the confines of the simulated worlds [for example Bruckman 95, Cherney 94 and Reid 94].

WOO: A WOO (Web MOO) is a MU* augmented by web pages for each of the spaces. Although movement among these spaces can occur with the graphical environment of the web pages, communication with others, as well as other actions, must occur in a text-based Telnet session running alongside a web browser.

The addition of web graphics allows those who construct the spaces to give others a clearer picture of those spaces, and allows members to illustrate their environment, but graphics cannot generally be used in conversation, though graphics in the environment might be 'pointed out' in conversation.

Internet Pager: Pagers allow a short text message to be sent to an individual specified by an indirect address. If the individual is currently on-line the message arrives immediately, indicating it's presence, otherwise it is queued until they reconnect to the Internet.

ICQ: As with many products, *ICQ* supports the functionality of a number of categories of tools, in particular multi-participant Talk and Email. It differs from other such tools in that an indirect address, an *ICQ number*, is used to locate other participants.

Agora: Agora [Long & Baecker 97] is designed to sit within the content of a web page. Like *Virtual Places* and some web chat systems, it shows who is browsing the same information and allows communication with them. In addition to real-time text communication, *Agora* supports a number of asynchronous methods of communication; a single threaded newsgroup, a person-to-person mail system, a history of recent visitors, and a persistent user profile that can be read by others.

Bulletin Boards: Bulletin boards are an interesting special case of application types. Most bulletin board systems, whether designed to run on the Internet, or to be accessed through local dial-up, support the features of IRC, email and newsgroups. However, they limit use to members of the particular board, thus creating a constrained user base.

By providing a broad range of tools (though all text based) to a limited set of users, bulletin boards are often able to support a long-standing community.

Conclusion

The number of tools available for online communication is ever increasing. The taxonomy of tools given here captures most of the major categories of the tools in use as of this writing. In researching the communities supported by these tools, the properties that make each tool different need to be considered. In addition, it is important to note the similarities between the tools, so that social phenomena observed in one tool might be extended to other tools. The properties of the medium exert a strong force on the character of the communities it hosts. The means of communication initiation, the conversational media, the style of interaction and the constraints placed on membership are important factors to consider when attempting to explain online behaviour. The role of dialog history and the differences between store-and-forward and real-time interactions are pivotal in the initiation of new members into online groups.

References

[Bruckman 95] Bruckman, A. and Resnick, M. (1995). Virtual Professional Community, Constructionism and Professional Community. *Convergence*, 1:1. http://asb.www.media.mit.edu/people/asb/convergence.html

[Cherny 94] Cherny, L. (1994). Gender Differences in Text-Based Virtual Reality. *Proceedings of the Berkeley Conference on Women and Language*, Berkeley, CA. ftp://parcftp.xerox.com/pub/MOO/papers/Gender.txt

[Damer et. al. 96] Damer, B., Kekenes, C. and Hoffman, T. (1996) Inhabited Digital Spaces. *Proceedings of the CHI '96, Conference Companion*, Association of Computing Machinery, New York, NY. Pp. 9-10.

[Donath & Robertson 94] Donath, J. and Robertson, N. (1994). The Sociable Web. *Proceedings of the Second International WWW Conference*. Chicago, IL. http://judith.www.media.mit.edu/SocialWeb/SociableWeb.html

[Long & Baecker 97] Long, B. and Baecker, R. (1997). Agora: Shaping Social Space from Information Space. *Proceedings of WebNet '97*, Association for the Advancement of Computing in Education, Charlottesville, VA. http://www.dgp.utoronto.ca/people/byron/webnet/Agora.html

[Morningstar 91] Morningstar, C. (1991). Settlements on the Electronic Frontier. Talk presented at IEEE CompCon '91. http://www.communities.com/papers/settlmnt.html

[Morningstar & Farmer 90] Morningstar, C. and Farmer, F. R. (1990). The Lessons of Lucasfilm's Habitat. In Benedikt, M. (ed.) *Cyberspace: First Steps*. Cambridge, Massachusetts: MIT Press. Pp. 273-301. http://www.communities.com/paper/lessons.html

[Reid 91] Reid, E. (1991) Electropolis: Communication and Community On Internet Relay Chat. Honours Thesis, Department of History, University of Melbourne. http://www.ee.mu.oz.au/papers/emr/electropolis.html

[Reid 94] Reid, E. (1994). Cultural Formations in Text-Based Virtual Realities. A Thesis submitted in fulfillment of the requirements for the degree of Master of Arts, Cultural Studies Program, Department of English, University of Melbourne. http://www.ee.mu.oz.au/papers/emr/cult-form.html

[Rheingold 93] Rheingold, H. (1993). *The Virtual Community: Homesteading on the Electronic Frontier*. Reading, Massachusetts: Addison-Wesley. http://www.well.com/user/hlr/vcbook/index.html

[Scardamalia & Berieter 91] Scardamalia, M. and Bereiter, C. (1991) Higher Levels of Agency for Children in Knowledge Building: A Challenge for the Design of New Knowledge Media. *The Journal of the Learning Sciences*, 1(1). Pp. 37-68.

[Turkle 95] Turkle, S. (1995). *Life on the Screen: Identity in the Age of the Internet*. New York, NY: Simon & Schuster.

Student Roles on the World Wide Web: Scholar, Researcher, Beggar, or Thief

Stephanie Ludi
Educational Media & Computers
Arizona State University, USA
sludi@asu.edu

Abstract

In recent years, educators have embraced the Internet and the World Wide Web in particular as the vehicle for change in education. The Web's hypermedia structure and diversity is viewed as a tool for exploration, creativity, and a means for interaction in the global community. Students and teachers use information and multimedia elements gathered from the Web to create written reports, multimedia presentations, and even other Web pages. These opportunities have a price attached as some students access inappropriate material or drown in a sea of information. In addition, many teachers still aren't trained in integrating the Web into the classroom. The resulting ignorance and naiveté about the Web has allowed many teachers and students to misuse the Web. This paper explores the various roles that student's play on the Web, how many educators advocate these roles, and what needs to be done to maximize on-line educational activities for students.

Introduction

Within the past couple of years, educators have embraced the Internet and the World Wide Web in particular as the vehicle for change in education. The hypermedia structure and diversity of the Web is viewed as a tool for constructivist curricula under the guise of preparing students to interact in the global community and tap into this information resource. Students and teachers have used information and multimedia elements gathered from the Web to create written reports, multimedia presentations, and even other Web pages. Communication with "experts" in topics and subjects studied in class is also encouraged.

These opportunities have a price attached. Many schools use the Web without an Acceptable Use Policy, the result is often a mixture of luck, chaos, and denial. Some students have accessed educationally inappropriate materials when unsupervised and left to surf the Web without an educational task to accomplish. Also, the Web is portrayed in the mass media as being everything from an encyclopedia to an information warehouse. This misrepresentation has lead to false impressions and false expectations [Soloway and Wallace, 1997]. In addition, many teachers still aren't trained in using and integrating the Web into the classroom. The resulting ignorance and naiveté about the Web has allowed many teachers and students to misuse the Web.

The anonymity and uncensored nature of the Internet/WWW has brought out the best and the worst in people, and students are no exception. This paper will explore the various roles that student's play on the Web, how many educators advocate these roles, and what needs to be done to maximize the on-line educational activities for students.

Future Consumers Versus Future Builders

Many educators have gotten the impression that students need to be computer literate and be problem-solvers in order to be prepared for tomorrow's workplace [McLain, 1997] and [Hawisher and Selfe, 1997]. This impression has been the result of pressure from business, parents, and the government. In order to truly prepare out children for the future, their educational opportunities must reflect the role that we want (and need) our children to have. Will our children passively use the ideas of others without any analysis or reflection, or will they integrate ideas into their own thoughts to build and create? Essentially we must ask ourselves *Do we see our children as future, passive consumers or as future, active builders?* The question is crucial, because the answer should guide the type and quality of the educational opportunities conducted in school. The answer may seem obvious - children need to be the active builders of tomorrow - but many Web projects and tasks undertaken by students does not reflect the role we want them to partake.

When student's download graphics, video, sounds, and text from Web sites to create a multimedia presentation, they are taking the role of consumer-in-training. All to often I have seen student multimedia presentations be a brief montage of other people's work - with no thought and no analysis. Worse yet, citing on-line resources is not stressed enough at school. In order to have students take on the role of scholar, they must analyze the information they are reading, listening, and viewing, and present their analysis in their projects [Mankato Schools, 1996]. Bibliographies should also be required for all (multimedia) projects.

Research on the Web: The Information Pyramid

The Web is a dynamic information resource that is often regarded with more credibility than it deserves. Since material can be published on-line regardless of the content, careful analysis of on-line materials must be conducted by students during the course of their schoolwork [McLain, 1997]. Such analysis and critical thinking enables students to become researchers and scholars.

When students navigate the Web for reference material, they should ask themselves several questions regarding the source of material [Grassian, 1997]. These questions can be categorized as General Purpose, The Author, and The Information. The following questions provide a framework for students to use to evaluate the quality and appropriateness of Web material.

General Purpose:
Is the purpose of the Web page clear to the reader?
Do the links clarify the purpose or supplement the objectives of the Web page?
Who is the audience for the information on the Web page?

The Author:
Who/What group is the source? Are they credible? Biased?
Does the author have any expertise in the topic? Is the expertise verifiable?
Is the Web page sponsored by a group/organization? Are they biased?

The Information:
Is the information 'refereed'? (i.e. Was the information evaluated and approved by an editor?)
Is the information persuasive or expository in nature?
Does the information contain facts or anecdotes?
Is the information complete and accurate?
Is the information presented in the Web page valuable in comparison with that of other information sources?
Is the information current in respect to the topic/issue being researched?

The answers to these questions will enable students to filter out the information that can be used in their projects. Also, it provides a good exercise in sorting through various types of information quality, sources, rationale, and bias as well as content.

Besides filtering information, students should use a strategy when working on projects. Because the Web is immense in the quality and quantity of available information, some structure must be imposed to prevent students from wondering aimlessly on-line. The Information Pyramid [Fig. 1] presents a strategy for students to follow while completing a project. Besides providing a step-by-step process, the diagram also illustrates how the information available on the topic decreases in quantity as the project becomes more focused. As the project progresses, the student should be conducting directed searches.

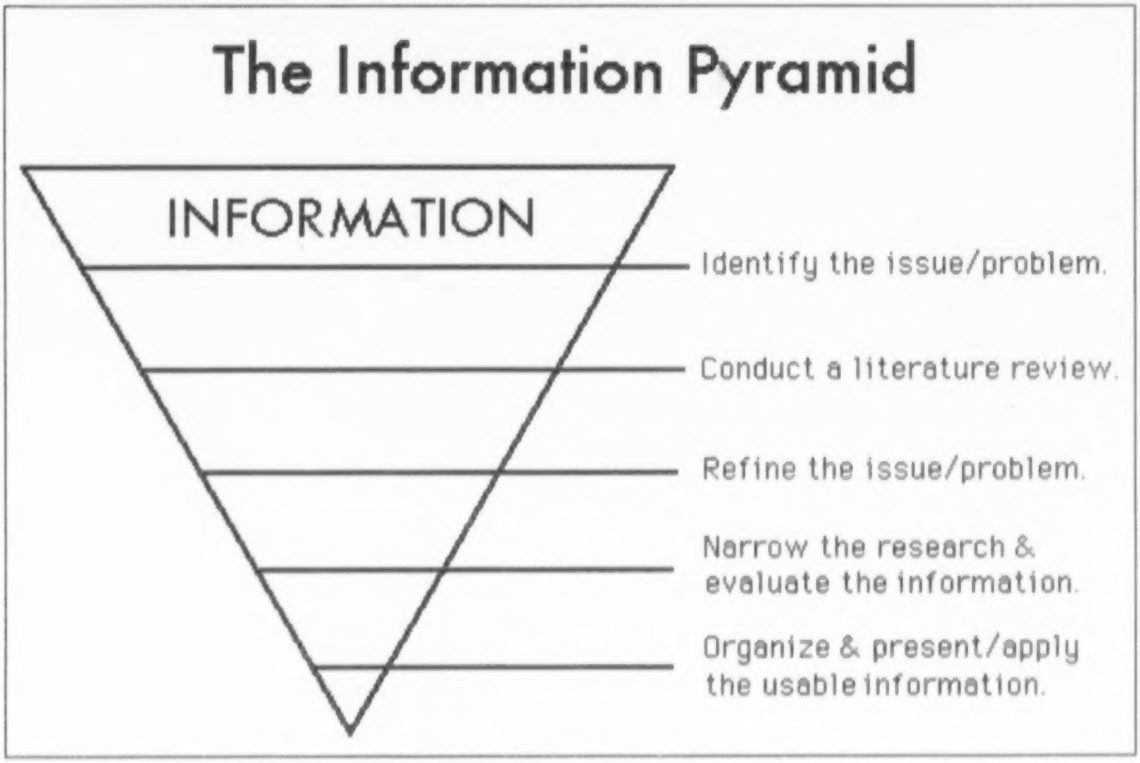

The Information Pyramid
INFORMATION
Identify the issue/problem.
Conduct a literature review.
Refine the issue/problem.
Narrow the research & evaluate the information.
Organize & present/apply the usable information.

Figure 1: The Information Pyramid

Fair Use & Copyright

The appeal of the World Wide Web is the integration of graphics, video, and sound with text into a format that can be accessed and downloaded. As a result, many students are encouraged to take information and multimedia files to use in multimedia presentations and papers. Students must learn how to cite the on-line material that they use, even if it is in a multimedia presentation. In a presentation, a screen can be created with a bibliography of the materials used. Citation methods, including the APA Style [Georgia Southern University, 1995], exist for referencing on-line materials. Expecting students to cite on-line materials in their projects enables the educator to gauge how much of a project's material is from on-line sources and who students are referencing.

Due to copyright and fair use issues that remained unresolved for years, the Fair Use Guidelines for Educational Multimedia were designed through a consortium of educational, publishing, and entertainment organizations [Penn State Libraries, 1996]. The purpose of these guidelines is to designate how copyrighted multimedia materials can be used in student and teacher projects, without getting consent from the copyright owner. While the guidelines may appear strict, they are not. The guidelines protect the owners of copyrighted materials from unauthorized use. In addition the guidelines promote the conservative use of copyrighted works in student projects and an increase in student-generated text and multimedia. By using the

Fair Use Guidelines, adhering to copyright laws, and citing on-line materials students will be trained to become researchers rather than thieves when on-line materials are used in projects.

Begging for Answers

In order to promote interaction in the Internet community, some students are encouraged to communicate with people with insight into topics being studied. At other times, students undertake such communication on their own initiative. While many students conduct themselves appropriately, too many students expect their homework to be done for them. When such students take on the role of beggar, they send email to other on-line citizens asking for the answers to questions that they should be researching. Since some Web browsers allow interaction in Usenet newsgroups, students have access to an audience of content experts. In order to curb these academically destructive habits, students' e-mail should be monitored and Acceptable Use Policies must contain a provision that expects students to do their own work. Such expectations must also be enforced.

Summary

While this paper may appear anti-Internet and anti-Web, it is not. Instead, the theme of this paper is that educators must ask themselves "What role in society should our students aspire to?" Pressure exists for teachers to prepare students for the technical world of the future, both in the workplace and society as a whole. If students are to ascend to positions of active leadership and become the builders and thinkers of the future, they must be trained and expected:

to critically analyze information.
to integrate this information with their own ideas and hypotheses to form a coherent argument or theme.
to follow a research process, such as the Information Pyramid, in order to become familiar with the benefits of planning and strategy in a project's evolution.
to balance the proportions of text and multimedia elements gathered from the Web in their projects according to the Fair Use Guidelines.
to cite information and multimedia elements gathered from the Web in all projects.
to use the Internet and Web responsibly.
to do their own work.

With these suggestions, students will be prepared to think critically, organize their thoughts, and act responsibly. Performing these duties with information from the Web in school will provide practice for gathering information in other contexts in the future.

References

[Butler, 1996]. Butler, B. (1996). Information literacy issues. [Online]. Available: http://www.gsia.cmu.edu/bb26/papers/education/iuc96/info-literacy.html

[Georgia Southern University, 1995]. Georgia Southern University Zach S. Henderson Library (1995). Preparing Bibliographic Citations APA Style [Online]. Available: http://www2.gasou.edu/library/broch_ref/bibliog.html

[Grassian, 1997]. Grassian, E. (1997). Thinking Critically about the world wide web resources. [Online]. Available: http://www.ucla.edu/campus/computing/bruinonline/trainers/critical.html

[Hancock, 1993]. Hancock, E. (1993). Information literacy [Online]. Available: http://www.ascd.org/services/eric/ericinf.html

[Hawisher and Selfe, 1997]. Hawisher, G., & Selfe, C. (1997). *Literacy, technology, and society: confronting the issues*. Upper Saddle River, NJ: Prentice Hall Inc.

[Mankato Schools, 1996]. Mankato Schools (1996). Information Literacy Curriculum Guideline [Online]. Available: http://www.isd77.k12.mn.us/resources/infocurr/infolit.html

[McLain, 1997]. McLain, T. (1997). *Internet homework helper*. Upper Saddle River, NF: Prentice Hall Inc.

[Penn State Libraries, 1996]. Penn State Libraries (1996). Fair Use Guidelines for Educational Multimedia and Related Documents and Links [Online]. Available: http://www.libraries.psu.edu/avs/fairuse/default.html

[Soloway and Wallace, 1997]. Soloway, E., & Wallace, R. (1997). Does the internet support student inquiry? Don't ask. *Communications of the ACM*, May, 40, 5, 11-16.

[Stanford University, 1997]. Stanford University (1997). SUL: Copyright & Fair Use [Online]. Available: http://fairuse.stanford.edu/

DIVERSITY IN ONLINE COURSES

Dr. Robin Mason
The Institute of Educational Technology
The Open University
England
r.d.mason@open.ac.uk

Abstract
This paper discusses the concept of diversity in the context of an online staff development course at the UK Open University. In addition to FirstClass computer conferencing and Web content, RealAudio was used in both synchronous and asynchronous modes. The purpose of its use was to add diversity and focus, which would maintain active participation until the end of the course.
After a description of the aim and components of the course, an evaluation of its effectiveness is made, and conclusions are drawn about an appropriate level of media diversity.

It is my experience of teaching and learning that people welcome a 'varied diet' when taking courses, in order to maintain interest in the subject matter, to motivate everyone to continue to the end of the course and to appeal to different learning styles. By a varied diet, I am thinking of the following elements: both synchronous and asynchronous learning modes; the use of multiple media such as text, audio and video or visual components; and a variety of ways of interacting with the course ideas from individual study and direct feedback from the teacher to discussion with other learners and collaborative activities (Laurillard, 1993). This diversity of course components used to be the prerogative of campus-based, face-to-face teaching. Technological developments have now made this diversity possible for courses taught at a distance.

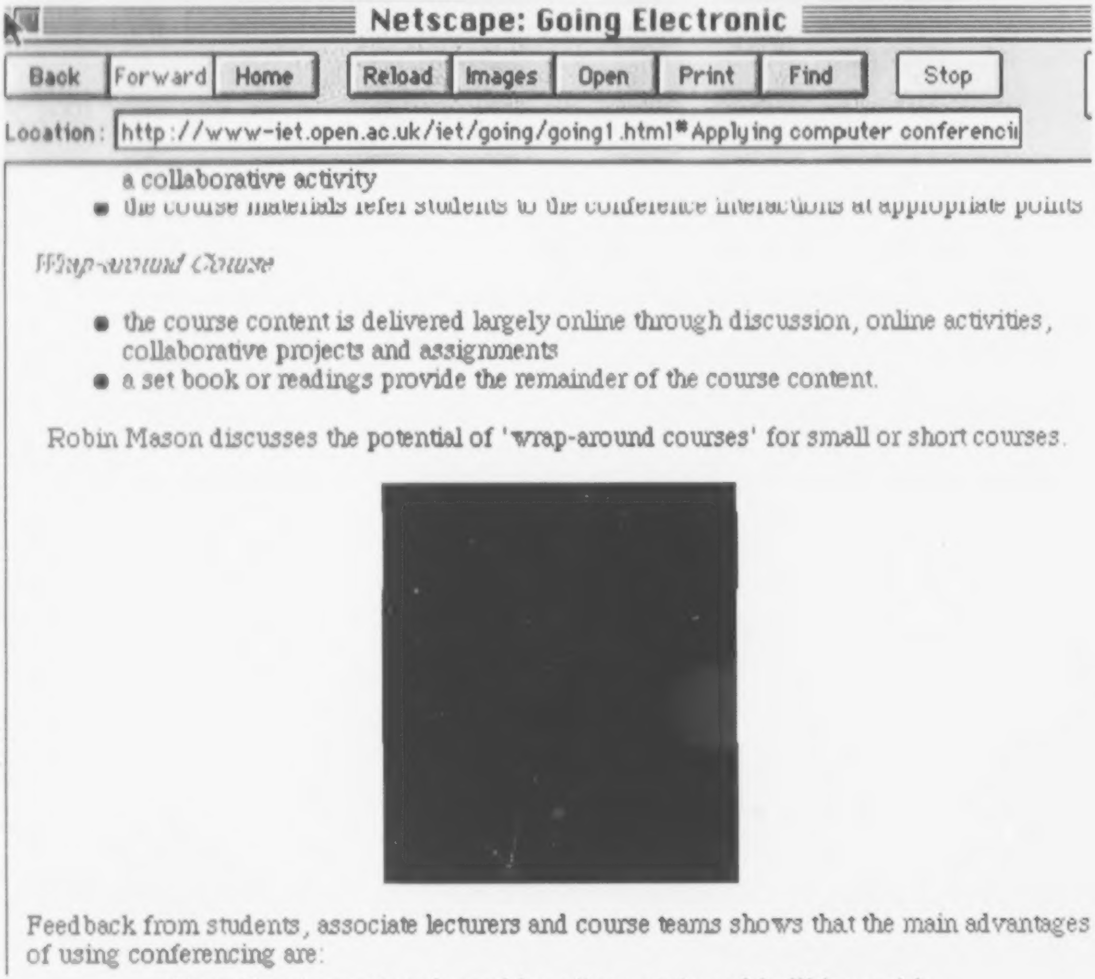
Just as technologies can and often are used 'for their own sake', so diversity in the form of multiple course components can be overdone. Students studying at a distance do not react favourably to a course which requires them to master many different media (several computer software programs, telecommunications and multimedia resources for example). They also complain when the course materials direct them, within a couple of hours of study time, to readings, then to computer activities, then back to the course materials and to a video before they can proceed with the next component (Morgan, 1989). In short, if some diversity is good, more is not necessarily better!

On the other hand, courses which are delivered primarily through one medium – say asynchronous computer conferencing – tend to have a falling off of participation. Asynchronous text messaging is a very flexible learning medium for fitting in to busy schedules, and encourages reflection and writing skills in the language of the discipline. But it is less powerful than real-time interaction as a means of motivating participants and maintaining commitment to completing the course (Mason, 1994).

After experimenting with many ways of creating a vibrant learning environment on an online asynchronous course, I came to the conclusion that real time events were the missing ingredient. This original asynchronous course, called Teaching and Learning Online, uses FirstClass computer conferencing to give trainers and educationalists hands-on experience of how to design and moderate an online course. While the 50 or so participants (spread around the UK and abroad) always start enthusiastically and work through the first two stages of the course with high levels of participation, there is always a marked dropping off of interest and interaction thereafter, despite every effort on our part as tutors and many refinements to the content of the second half of the course. Feedback from participants indicated that, as a professional updating course fitted into the spare moments around many other commitments, taking part in the course simply 'fell off the end of their list of things to do in the day' (Wegerif, 1995).

GOING ELECTRONIC

I had the opportunity to re-think the course for staff of my own institution, the Open University (OU). As participants on the course had access to technical support and relatively powerful computers, I decided to try an online real time event at the end of the course, and to use it as the focus for drawing together all the elements of the course. In addition, I used the same technology (RealAudio) to 'annotate' the Web pages which carried the content of the course. My aim in writing the Web content was to draw together institutional experience about how to design, instigate and manage the online component of courses within the OU. Expertise in these various areas is spread widely throughout the organisation, ranging from the chair of a course team, to the course manager, tutors, administrators and systems support staff. In order to capture some of this expertise, I asked representatives of these various areas to make recordings in which I questioned them about their particular experiences of using FirstClass. These recordings were edited into 'sound bites' ranging from 2 to 5 minutes and added as audio clips (with the photograph of the speaker) to appropriate places in the Web pages. The Web pages remain as an institutional resource in addition to their use for the course. Figure One shows an extract from the materials.



http://www-iet.open.ac.uk/iet/going/going1.html#Applying computer conferenci

a collaborative activity

- the course materials refer students to the conference interactions at appropriate points

Wrap-around Course

- the course content is delivered largely online through discussion, online activities, collaborative projects and assignments
- a set book or readings provide the remainder of the course content.

Robin Mason discusses the potential of 'wrap-around courses' for small or short courses.

Feedback from students, associate lecturers and course teams shows that the main advantages of using conferencing are:

Figure One: A snap shot of one of the audio annotations of the Web materials

The only paper component of the course was the following course outline which gave participants an overview of what they would be doing:

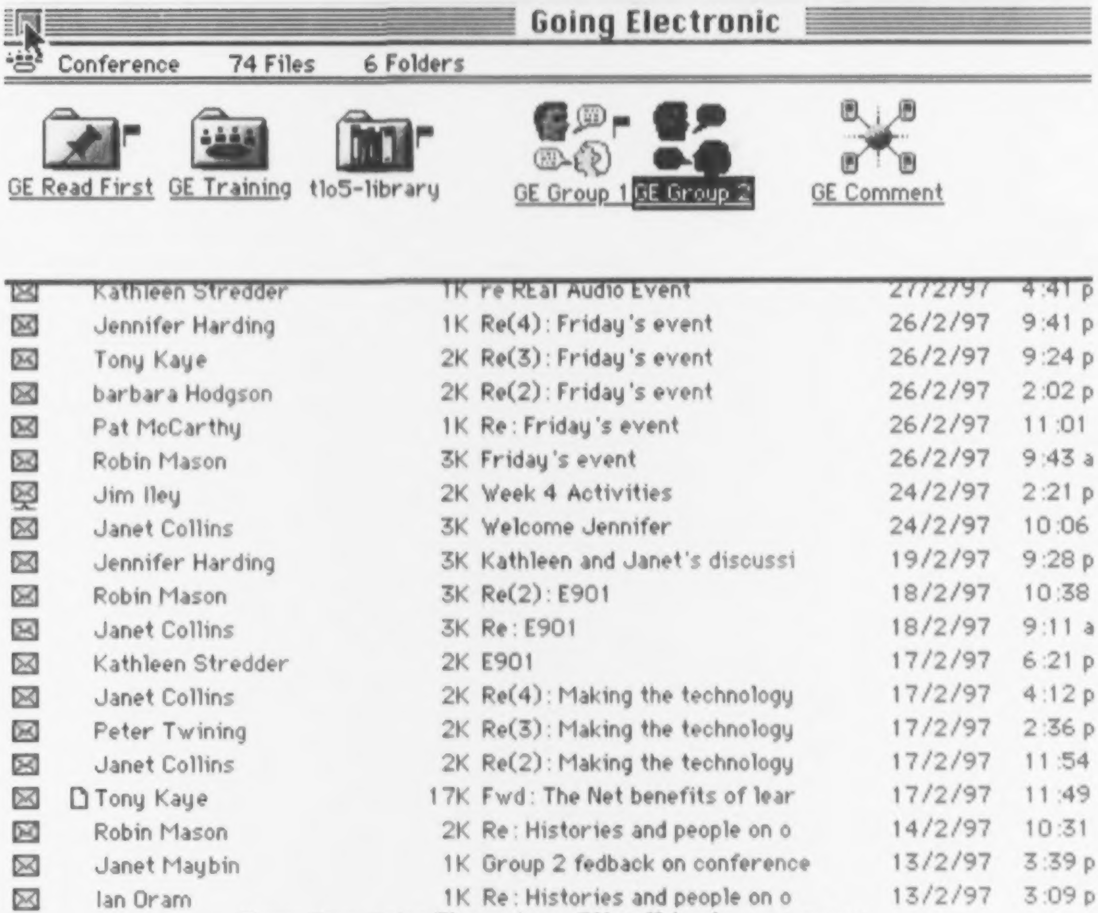
BEST COPY AVAILABLE

> The course will begin with a half-day face-to-face meeting at the Stony Stratford training centre. This will familiarise you with FirstClass, the Web/Real Audio, and with the other members of your group. However, the rest of the course is delivered online. Obviously you will need to have a computer which runs FirstClass and Netscape, preferably, but not necessarily, on your desk. Audio Visual will co-ordinate the RealAudio event and in principle, you can take part from the regional offices, or from various sites around Walton Hall. The course will last for 4 weeks with the following agenda:
> Week One: Face-to-face meeting, training exercises in using FirstClass, online interaction with other members of your group, working through Web materials
>
> Week Two and Three: Online debate about issues raised in the Web pages in which you will be given a role to play such as proposer or opposer of the motion, moderator of the discussion, commentor, summariser.
>
> Week Four: Working in small groups online to prepare a group presentation, and participating in the RealAudio event.
>
> I anticipate that the whole course should take between 20 and 40 hours, depending on your previous familiarity with the various media, and your commitment to all aspects of the course.

The course interactions took place on FirstClass because this is the system currently in large scale use at the OU and the course was about this use. The Web could equally well have been used – in fact, perhaps better, in that the online discussions could have been linked with the associated Web pages. Figure Two shows the opening screen of the FirstClass course area.

EVALUATION

There are three elements to the evaluation of the course:
•did the live event succeed as a motivator to keep participants engaged in the course?
•did the audio clips in the Web pages help to provide diversity on the course?
•did the combination of Web-delivered content and online interaction succeed as a useful learning environment?



Kathleen Stredder	1K	re REal Audio Event	27/2/97	4:41 p
Jennifer Harding	1K	Re(4): Friday's event	26/2/97	9:41 p
Tony Kaye	2K	Re(3): Friday's event	26/2/97	9:24 p
barbara Hodgson	2K	Re(2): Friday's event	26/2/97	2:02 p
Pat McCarthy	1K	Re: Friday's event	26/2/97	11:01
Robin Mason	3K	Friday's event	26/2/97	9:43 a
Jim Iley	2K	Week 4 Activities	24/2/97	2:21 p
Janet Collins	3K	Welcome Jennifer	24/2/97	10:06
Jennifer Harding	3K	Kathleen and Janet's discussi	19/2/97	9:28 p
Robin Mason	3K	Re(2): E901	18/2/97	10:38
Janet Collins	3K	Re: E901	18/2/97	9:11 a
Kathleen Stredder	2K	E901	17/2/97	6:21 p
Janet Collins	2K	Re(4): Making the technology	17/2/97	4:12 p
Peter Twining	2K	Re(3): Making the technology	17/2/97	2:36 p
Janet Collins	2K	Re(2): Making the technology	17/2/97	11:54
Tony Kaye	17K	Fwd: The Net benefits of lear	17/2/97	11:49
Robin Mason	2K	Re: Histories and people on o	14/2/97	10:31
Janet Maybin	1K	Group 2 fedback on conference	13/2/97	3:39 p
Ian Oram	1K	Re: Histories and people on o	13/2/97	3:09 p

Figure Two: Going Electronic, an OU staff development course

RealAudio Event

The course concluded with a one hour 'audiographic' event, in which participants went to particular locations on campus, or used the machine on their desk if it was powerful enough and they had already downloaded the RealAudio software. Participants heard three short presentations summarising the online interactions, and followed a Netscape screen of prepared overheads. They could send an email question or comment at any time and could also view all those submitted by the other participants. The final half hour of the event was an informal discussion by the three presenters about the comments submitted by email during the event. Both the presentations and the comments continue to be viewable and audible from the same site as an asynchronous resource.

Without any prompting, I received the following feedback from one of the participants who had managed to set up RealAudio on her own machine:

> I had a meeting at 10 so was too late to hear Peter [the first presenter]. So thank you for the brilliant idea of a replay and I could also stop the tape and make notes. You were riveting (perhaps you had your eyes closed and so could see us all in front of you). Peter was also excellent - I liked his boings. [telling the audience to move to the next overhead!]. The value of this event seems to me to lie in how well the presenters summed up. It doesn't matter that more questions are raised than answered. The questions and observations from you and Peter have given me a sense of a conclusion to the course and the beginning of my investigations of real conferencing. The event has also provided me with more motivation and given clarity to the issues to be solved.
> Many thanks, again

And from another participant who was unable to attend the live event:

> I'm afraid I won't have any time on Friday to participate in the RealAudio event....I'm very disappointed as I should like to have learnt how to set up a real audio page on the web (and also of course to have properly finished the GE experience - which I have thoroughly enjoyed....so thank you!). I feel rather frustrated, because the course will fizzle out for me and there will be no proper endings, either to the Group 1 conference or to GE.....oh well.....I have learnt a lot, and am thoroughly converted to the use of conferences in course provision....thank you very much for your guidance.

As tutor, I also felt that the live event concluded the course in a more positive way than I have ever experienced with previous online courses. There was a greater sense of momentum building up to the event, and, although not all of the staff who 'signed up' for the course actually participated, those who did remained active until the end.

Audio clips

The fact that quite a few participants did not make the effort to get to a machine where they could hear the audio clips indicates that perhaps this was one diversity too many! I did receive some very valuable feedback from one participant about how the sound bites should differ from the Web text materials:

> Having experienced the elation of actually managing to listen to the audio sequences, here are some comments about how I reacted to them and how useful I found them.
>
> 1 Applying computer conferencing to the OU context
> I liked the Web notes; they gave a succinct point by point resumé of the benefits and uses of CMC and CML. However, I thought the audio clips did not give added value except (i) in Gary's case to mention the use of a practise conference and of maintaining a rich and lively environment for the students, (ii) in Robin's case the mean's by which courses can maintain up-to-datedness and (iii) in Gilly's case to use the medium holding academic and student support objectives to the fore.
>
> 2 Integrating conferences with other course components
> Again, I found these notes to be useful, but wasn't particularly interested in the information being given by the audio clip.....however, the clip did begin to give me a feel for how the medium could be put to use within a course context.
>
> 3 Preparing Associate Lecturers
> Once more the notes provided valuable points to be aware of in designing a course that will involve conferencing. Gilly's comments on training Associate Lecturers were quite useful but too general....I wanted to know more about the content of the training course.
> This is where I could see how to use the audio clip could provide a different medium for conveying information.....the notes contain some very detailed points and I was beginning to sag a bit at reading them...but I could listen to someone telling me some more specific points....However, I realise some lighter relief from the detail also helps....it just seems to me that all the clips so far have tended to the light relief side of things.
> Nick's first clip reiterated some of the points in the notes, but it was worth hearing how they tried to tackle some of the problems...unfortunately the clip finished just as he was about to describe how the course team have tried to overcome the issue of time management...I'd liked to have heard this.
> His second clip was spot on.....I especially found his warnings helpful...not to overuse the facility to post stop presses etc., for the course team to beware of the disproportionate weight given to their opinion.
>
> 4 Technical considerations
> Now this is where I found the audio clips coming into their own. Apart from Pete Thomas' contribution, which, because of its historical bias and therefore lack of relevant information (even in a generalist sense), I found of little value, all the clips here gave information that was ONLY

accessible via audio, i.e. they were not repeating the text, nor were they of a purely general nature. I paid much more attention to these than I had to any of the previous clips.

5 Student considerations
To my surprise, given my comments above about using audio for generalist contributions, I found Robin's clip worked well. This seems just the place to use such contributions....at the beginning as an intro.

6 Conference structures and conference futures
Yet again, I found the first contribution, by Nick, of little value because it simply repeated the text. Tina's contribution mixed information presented in the text with new ideas...OK, but I found it difficult to maintain interest throughout. Nick's second contribution was much better, because it introduced new information only through the audio clip, and Pete's contribution on delivering a computer conferenced tutorial in real time was again valuable because the info could not be found in the text. Gary's contribution lacked interest because it was simply wallpaper, and Nick's third clip simply reiterated the text on critical mass....a process that you might have gathered by now I have come to dislike. However, Ben's discourse on the Off Line Reader in Conferencing Futures was excellent, again because she was passing on important detailed comment about the merits (or otherwise) of the system and because this also could not be found in the text. In fact, I would go so far as to say that I found this the most useful of all the audio clips on these pages.

I will certainly be more careful in any future audio annotations I make, that the content of the audio develops, enlivens or details the information in the Web text. Merely confirming what the text says, even if the confirmation is by a known expert, does not justify the use of another medium. While I was aiming for a 'fireside chat' feeling to the audio clips, I can see that the chat needs more scripting than I realised first time around.

Web Content and Online Interaction

In Stage Two of the course, where the group was divided in half, each with a different topic to discuss, participants were invited to use the Web materials as a foundation for their arguments for or against the particular question posed as the focus of discussion. In one of the groups, each participant was assigned a specific role (and several of these required the participant to refer to the Web materials); in the other group, participants were free to interact in whatever way they chose. As our previous experience with these two extremes has confirmed time and time again, a structured environment works much better: people get on with the task they have been assigned and most members of the group are active participants (Mason, 1997). Their concern about letting the group down overcomes their inhibitions about participating and the designation of a role helps them to focus on a particular form of input.

The 'free-for-all' method, on the other hand, often leads to one-sided discussions between a few enthusiastic participants. Others soon feel they don't want to intrude in the two or three way conversation and they either watch or drop out.

In the case of Going Electronic, the 'structured' group did make many references to the Web materials, often quoting or paraphrasing extracts, while the unstructured group never referred to them. This is not to say that their discussions were irrelevant, but merely that, from the point of view of a course aiming to convey a certain body of information, a tight framework needs to be in place to integrate the content with the discussion of it.

CONCLUSION

I am satisfied myself that the experiment in using RealAudio was worthwhile; that is, worth the extra effort of staff in supporting and taking part in it. Novelty, I realise, may be playing a role in this perception. Online courses are not new for me, whereas designing audio clips and a RealAudio event certainly were. However, the greater commitment of many (though not all) participants up to the end of the course does confirm my hunch that real time events do add an important ingredient to online courses.

On such a short course as this (four weeks) I perhaps went overboard in the diversity offered to participants. Learning to use FirstClass, accessing Web pages, taking part in online discussions and attending a face-to-face event at the beginning and a RealAudio event at the end, was enough 'diversity' for a staff development course which was added onto all the usual job and domestic commitments. The technical barriers of getting RealAudio working on their desktop, or taking the time to go to a machine where the software was already installed, was one too many demands. However, the great thing about these technologies is that the resources and materials created for them continue to be available long after the formal course is finished.

REFERENCES

Laurillard, D. (1993). *Rethinking University Teaching*. Routledge, London.

Morgan, A. (1989). Home Computing Evaluation Project: Students' experiences of study. SRC Report No. 28, Institute of Educational Technology, The Open University.

Mason, R. (1994). *Using Communications Media in Open and Flexible Learning*. Kogan Page, London.

Mason, R. (1997). *Global Education.* Routledge, London. (in press)

Wegerif, R. (1995). Collaborative learning on TLO'94: creating an online community. CITE Report No. 212, Institute of Educational Technology, The Open University.

Note: As the Web materials and the RealAudio event were produced only for internal OU staff, the url to the Web site is not for public use.

Computer Mediated Communications: for Dabblers or Practitioners?

Jon Mason
Faculty of Education
j.mason@edfac.unimelb.edu.au

Matthew Riddle
Science Multimedia Teaching Unit
m.riddle@science.unimelb.edu.au

The University of Melbourne, Parkville, Victoria 3052, Australia

Abstract: With the current high interest right across the education sector in the various enabling communication tools characteristic of the Internet and Intranets, action research has almost become an essential requirement. The diminishing shelf-life of both technology and the knowledge which must accompany it deem that this is so. This has led, however, to what could be termed a clash of culture in higher education, a 'collision of timezones'. For as durable as 'traditional' academe has been, the signs of not keeping up are all too clear. Despite the rush into using the Web for what it seems to offer, real 'practitioners' in Computer Mediated Communications (CMC) are generally fewer than all the hype would suggest. 'Dabbling' as a means of keeping up has become elevated to an art form — knowing just enough to make commentary 'just-in-time'. Perhaps the nature of Web 'browsers' themselves has contributed to this syndrome, because browsing is, after all, what could be called the 'first level' of Web awareness. The upheaval in the sector, of course, is also due to a number of other factors — such as the contraction of public funding and increased competition due to globalisation. Nonetheless, CMC has an integral role in the transformation of higher education and its rigorous usage by practitioners focused on its application to teaching and learning indicates what may be a viable and durable mode of delivery and access into the next decades.

Introduction

CMC (and other 'multimedia' technologies) at the University of Melbourne (as in many other universities) is currently being used explicitly for its educational potential, for its capability to "transform teaching and learning". [Hart & Mason 1996, Taylor et al. 1996] Why has this come about? John Tiffin and Lalita Rajasingham [Tiffin & Rajasingham 1995] provide useful perspective in their description of education as communication supported by their argument that communication operates at multiple levels: intrapersonal, interpersonal, group, organisational, mass, and global:

> "Education systems are complex communications systems concerned with the transmission, storage and processing of information. Their purpose is to assist learners so that from being unable to deal with problems they become proficient problem solvers. This depends on communication networks that intermesh four related factors: learning, teaching, knowledge and problem. There appears to be a fractal dimension in that the network that intermeshes the four related factors can prove to be a node in a network at a higher level. Similarly, a processing node in a network can, at a lower level, prove to be a network. The existence of different levels in a communications system for learning allows learners to shift levels in the process of learning." [Tiffin & Rajasingham 1995]

Importantly, the digital domain brings to the world of communications both enhancements and deficits. Some technologies render earlier technologies obsolete while others extend the opportunities and functions for interaction. Thus, while the typewriter didn't replace the pen, the telephone replaced the telegraph, the wordprocessor has nearly done away with the typewriter whereas mobile telephony is an extension, not a replacement, of conventional telephony services. CMC clearly extends the domain of human communications but it is a technology yet to achieve more than a niche audience. CMC can embrace many forms of conventional communication: writing, speaking, listening, reading, watching, reflecting, posturing, gesturing, etc. — all these kinds of activity can take place successfully in CMC.

However, while this is not always the case, CMC (particularly in asynchronous mode) also offers major benefits such as the opportunity to reflect on discussions over long time frames and the ability to manipulate, collaborate and share substantial amounts of information, often in mixed data formats. For this reason, CMC has a uniqueness which differentiates it from other distributed communications systems such as broadcast (radio and television) and print media but it is our experience that novices often come to this new medium with expectations coloured by these other media. The mainstream 'media' are partly responsible for this because of their excessive portrayal of 'cyberspace' and 'virtual' communication as an 'other' realm, as some kind of substitute for real communication. Yes, the hype points to its mass arrival but its 'otherness' still remains. But do we think of holding a piece of plastic to our heads and listening and talking as 'virtual' communication or 'real' communication? Most of us certainly take telephone conversations for granted as normal daily human interaction!

There are many other examples of the transparecy of 'virtual' worlds in our daily lives: for example, where is the music that we listen to? Popular analysis suggests that it is merely a combination of 'organized sound' which has 'emotive expression' and can be pleasing to the senses. But it certainly can also facilitate abstract reverie. And *where* is it? Not just in the instruments. Not just in the sound waves. Not just in the concert hall. Not just in the sound system. Music possesses layers of structure and is rich in ambiguity. It can be emotively, cognitively, kinesthetically experienced and musical perception is a far more complex process than just auditory perception. [Serafine 1988]

Jo Ann Oravec takes this line of argument further and devotes a whole chapter to the "infinite variety of virtual entities" in which she says:

> "to an increasing extent, management in organizational contexts has become the management of virtual individuals and groups. Virtuality has become a common theme in American life, taking on connotations of the "imaginary," as well as the "designed" or "engineered": "virtual corporations" are created when corporations extend their spheres of influence." [Oravec 1996]

In our experience with CMC we have identified a number of 'levels' of activity, though in practice, there is a continuum of activity in CMC which ranges from browsing (what Tiffin et al. might characterise as communication on the 'intrapersonal' level) to extensive communications developed within virtual communities of practitioners. [Rheingold 1994, Norris 1996]

While we recognise that CMC has also entered the workplace, and clearly universities are also workplaces, our model is more suited to the educational context. CMC in the workplace context is usually referred to under a different acronym, CSCW, or Computer Supported Co-operative Work.

Background

'CMC' itself is often terminology used by educationists to describe the merging of telecommunications and computer technologies in providing new teaching and learning environments — new ways of human communication.

> "CMC describes the ways we humans use computer systems and networks to transfer, store, and retrieve information, but our emphasis is always on communication." [Berge & Collins 1996]

That's a fairly standard description. But what it doesn't describe is the inherent limited shelf-life of this technology, which is subject to rapid and constant change. Electronic texts might be more flexible and dynamic but they also lack a certain quality of endurance. As Ilana Snyder puts it in her book, *Hypertext:*

> "In contrast with traditional text, electronic writing depends upon an emergent technology which is still subject to transformation" [Snyder 1996]

Snyder offers further comment on the unique features of CMC, which as a medium is still currently dominated by text-based communication. "Writing with a computer not only blurs the line between thinking and writing but also shapes to some extent the ways in which we think." [Snyder 1996] With the computer, the window has an impact on the construction of texts, as do supporting icons and tools such as the mouse, the cursor, and the scroll bar. But also, "Readers of screen texts are denied some of the spatial-contextual cues to which readers of page texts have access" such as text length, page numbers and book thickness. She thus suggests there is a "grammar of the screen" which must be learned. [Snyder 1996]

Although coming from a different angle, this view finds resonance with that of Herbert Stahkle and James Nyce, who suggest that much of the rush toward Web-based delivery has been "ill-conceived" and that "the tendency has been to assume appropriateness" of the medium for more courses than are naturally suited to it." [Stahlke & Nyce 1996] However, they also argue:

> "By designing around asynchronous assumptions, distance learning can become a rich, varied, and highly effective modality, so much so, in fact, that the college or university may well see a need to design the on-campus educational experience modularly and asynchronously so that on-campus students can enjoy as rich an experience as the off-campus student."
> [Stahlke & Nyce 1996]

The Continuum of CMC

A common criticism of the increase in usage of electronic communication is that its form is destructive of the style of sustained polemic which has become possible in print, and that it instead encourages readers to move quickly from one idea to another, almost as though engaging in a form of word association.

> "Electronic communication can be passive, as with television watching, or interactive, as with computers. Contents, unless they are printed out (at which point they become part of the static order of print) are felt to be evanescent... The pace is rapid, driven by jump-cut increments, and the basic movement is laterally associative rather than vertically cumulative." [Birkets 1994]

Despite the so-called 'interactive' nature of the World Wide Web, simply browsing is often far less enaging (or truly interactive) than reading traditional, narrative-style, printed texts. However, criticisms such as this, which warn that new forms of electronic communication will destroy forever the ability of learners to interpret and form extended arguments, assume that new technologies will inevitably annihilate old. In addition, these arguments tend to characterise printed books as associated with logical constructions due to their linear progression. Printed words require active engagement and interpretation. Engagement in electronic communication is not bound by the order of the printed text — it is free-form and impermanent. [Birkets 1994]

Yet such a picture fails to capture the full experience of electronic communication because it attempts to judge new technologies by viewing them as a new means to the same end. If we examine the Internet, for example, as a new way to read the works of Shakespeare, or television as a replacement technology for learning history, it seems inevitable that we will find shortcomings.

Another way to conceive of CMC is to construct a continuum. Activities such as passive consumption of visual material are at one end, followed by browsing hypertext (or 'dabbling'), reading, observing engaging discussion, and basic interactions.

Towards the centre are activities which enable active discourse using asynchronous technologies such as bulletin boards and newsgroups or conferencing facilities, or synchronous technologies which may be text-based conferences (with or without graphical support), audio/videoconferencing, or indeed combinations of two or more. Included here are those technologies which are analagous to the writing and criticism of printed texts, with the addition of the increased capability of average 'readers' to publish their own works for public consumption (using the World Wide Web) as opposed to the relatively centralised concentration of power associated with the printed publishing process.

At the more developed end — what the authors regard as the practitioner's domain — are activities which appear to be difficult to categorise in terms of their analogy (or homology) to previous technologies. Firstly the combination of many interactive media-types which place relatively passive means of communication (or information retrieval) alongside those which are narrative, interactive, discursive, and or dynamic is something which is certainly new. Moreover, the more technologies such as the Internet and the World Wide Web evolve, it seems, the more the distinctions between these forms appear to become blurred, so that as we may seem to be consulting a computer we can be experiencing broadcasts, engaging in conversations, and publishing ideas simultaneously. An example is the arrival in 1996 of the so-called "push" technologies which deliver news articles or advertisements to the computer screen at pre-scheduled regular intervals. [Wired Magazine 1996].

Email could be considered a simple push technology when used in certain ways — such as listservs or for the infamous practice of email advertising — while television is undoubtedly the model for the push for push. What is different about the digital, media rich form of push is that the passive media exist alongside a large array of interactive communications technologies, making it possible to engage in a gameshow online (like the quiz game *You Don't Know Jack*) [HREF 2] rather than simply watching others on your television screen. It is this active engagement which places such new media at the more developed end of the continuum. It does not take a great deal of imagination to conceive of a push version of, say, David Attenborough's *Life On Earth* series by the BBC, which might allow discussion or deeper investigation of areas of interest. Further, it is our view that recognition and utilisation of the digital desktop as a natural multi-tasking environment typifies CMC practitioners.

Another important example at this end of the continuum is the emergence of 'groupware', or software environments which allow many people, typically workgroups, to work together on documents and projects.

> "CSCW applications' (or groupware's) emergence as a genre is significant in itself; it signals increasing levels of interest in the group or team as an organizational or social unit." [Oravec 1996]

Forums at The University of Melbourne

Teachers and learners are beginning to introduce CMC to the higher education experience at the University of Melbourne. On a forum hosted by the Faculty of Education, for example, students from a range of disciplines such as Arts, Early Childhood studies, or Botany can access online discussion areas set aside for them using the World Wide Web. [HREF 1]

These forums allow threaded, asynchronous discussions on topics relating to the students' coursework or related issues. They provide both public and private means of communication, and include access to synchronous text-based discussions, or 'chats'. Using this same environment, students can review lecture slides or access data referred to in discussions. They can publish their own documents quickly and easily using a 'newspaper' facility, and can be automatically notified of changes to the discussion by email if they desire.

At the same time, forums exist for the discussion of issues relating to online education and computer literacy for educators. A weekly electronic newsletter is sent by email to educators at the University and others (on and off campus) interested in issues relating to online education, encouraging discussion and involvement in communications forums. [Mason & Hart 1997]

Conclusion

It is important, we argue, that in order to understand and predict the use of CMC for education we recognise the continuum of CMC. In fact, the appropriate application of CMC technologies to education necessitates a full understanding and participation in activities towards the more sophisticated end of the continuum, due to the tendency of new media simply to be applied to old ways of teaching and learning, without really exploring their new potentials. This has the added advantage that traditional modes of communication are not simply expunged from the educational experience, but are added to, and hopefully enhanced.

References

[Berge & Collins 1996] Berge, Z. & Collins, M. (1996). *Computer Mediated Communication and the Online Classroom* New Jersey: Hampton Press.

[Birkets 1994] Birkets, S. (1994) *The Gutenberg Elegies: The Fate of Reading in an Electronic Age.* London: Faber & Faber.

[Hart & Mason 1996] Hart, G. & Mason, J. (eds.) (1996) *The Virtual University? — Symposium Proceedings and Case-studies* The University of Melbourne.

[HREF 1] EdFac Forum: http://forum.edfac.unimelb.edu.au

[HREF 2] Jack NetShow, Beserk: http://www.bezerk.com/netshow/

[Hunt 1996] Hunt, R. (1996) Collaborative Investigation Online: Eighteenth Century Literature Moves to the Computer Lab, in [Berge & Collins 1996] 93-110.

[Mason & Hart 1997] Mason, J. & Hart, G. (1997) Effective Use of Asynchronous Virtual Learning Communities. *Creative Collaboration in Virtual Communities 97,* University of Sydney. http://www.arch.usyd.edu.au/kcdc/conferences/VC97/

[Norris 1996] Norris, D. (1996) Transforming Higher Education — Unleashing the Power of Perpetual Learners in Hart, G. & Mason, J. (eds.) (1996) op. cit, pp. 17-21.

[Oravec 1996] Oravec, J.A. (1996) *Virtual Individuals, Virtual Groups — Human Dimensions of Groupware and Computer Networking.* New York: Cambridge University Press.

[Rheingold 1994] Rheingold, H. (1994) *The Virtual Community: Finding Connection in a Computerised World.* Minerva.

[Serafine 1988] Serafine, M. L. (1988) *Music as Cognition — The Development of Thought in Sound* New York, Columbia University Press.

[Snyder 1996] Snyder, I. (1996) *Hypertext — The Electronic Labyrinth.* Melbourne: Melbourne University Press.

[Stahlke & Nyce 1996] Stahlke, H.F. & Nyce, J.M. (1996) Reengineering Higher Education: Reinventing Teaching and Learning. *CAUSE/EFFECT* 19 (4) 44-51.

[Tiffin, J. & Rajasingham 1995] Tiffin, J. & Rajasingham, L. (1995) *In Search of the Virtual Class* Routledge: London.

[Taylor et al. 1996] Taylor, P., Lopez, L., & Quadrelli, C. (1996) Flexibility, Technology and Academic Practices: Tantalising Tales and Muddy Maps. *Evaluations & Investigations Program,* DEETYA: Commonwealth of Australia.

[Wired Magazine 1996] *Wired Magazine* 5.03 (1997), 'Kill Your Browser', March 1997, California, http://wwww.wired.com/wired/5.03/features/ff_push.html

Cost Reduction through an Intranet: The Paperless Office

by: David Matusow, NASA/Goddard Space Flight Center

with: Joe Sparmo and John Bristow

Abstract

With the continuing effort to reduce costs in all areas of government, the Flight Dynamics Division (FDD) at the National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (GSFC) has turned to the Internet as a way to make administrative functions more efficient and, therefore, more cost effective. Administrative functions, such as time and attendance collection, are a big part of any organization, and can either aid an organization towards its goals, or act as a parachute slowing or "dragging" an organization down. To minimize the "drag", the FDD has developed an Intranet, or Internal Web, which is accessible only to FDD related personnel. Employees log into the system through a user-id and password, which allows the system to serve only the appropriate personnel. Once logged into the system, the user has access to the many internal web pages that support the organization as well as what is called the *Paperless Office*. It is called the Paperless Office because the FDD has taken functions which have traditionally been paper based, and automated them into a collection of web-based tools. Some of the functions that these tools provide include:

- A leave slip
- A flexible time and attendance system
- A charge hours tracking program
- A flexible work schedule management tool
- A conference room scheduler
- Many standard Government forms

This paper will explain these systems in detail, and show how using this Paperless Office system has brought significant process improvement to the FDD resulting in measurable cost reduction.

Contents

1. What is the problem?
2. What we decided to do about the problem
3. Details of the Paperless Office
4. Conclusions

1.0 What is the problem?

Inherent in any organization is overhead. Most organizations deal with it. Many try to reduce it. Some just carry it as baggage. One of the biggest types of overhead is administrative functions. These functions are at the heart of every organization, and the more transparent they are, the smoother the operation and (usually), the happier the people.

Some examples of the problems with typical administrative functions are:

i. They are expensive. You need people to manage it, process it, maintain it, and people to train others in how to use it.

ii. They lower productivity. People have to spend significant amounts of time filling out paper forms.

iii. They are not user friendly. People have to remember job charge numbers, or know who to contact about a form which needs to be signed by someone whom they don't know.

iv. Most electronic COTS software that handle this type of information is expensive, and not platform independent.

These problems can be overcome, and in the NASA/GSFC Flight Dynamics Division, they have been greatly reduced by the Paperless Office system, which resides on the FDD Intranet.

2.0 What we decided to do about the problem

To reduce costs and create an overall better scheme for handling administrative duties, the FDD created the Paperless Office. This is a toolset, which allows the employees of the division to access most forms and information over an Intranet, without the user of paper. The creators of the Paperless Office decided to exploit the Internet, because it provided a mechanism to allow platform independent access to all forms, which were written in the Hypertext Markup Language (HTML) and the Practical Extraction and Report Language (PERL). This was necessary since some of the FDD employees only had access to workstations, while others had Macs and still others had PCs.

3.0 Details of the Paperless Office

As stated earlier, the Paperless Office is a toolset of Common Gateway Interface (CGI) scripts, written in PERL and HTML. These scripts allow the employees of the FDD to handle such administrative tasks as filling out a leave slip, completing their time and attendance record, recording and modifying their flexible work scheduler, filling out Government forms, and scheduling meetings.

3.1 Databases

The scripts were created to operate by accessing flat-file, ASCII databases. Simple ASCII databases were chosen to cut costs, while providing for platform independence and flexibility. These databases contain information about the organization. For example, there is a database of employees in the different branches within the division. Another database contains Job Order Numbers (JONS) and an acronym to go along with that number. These databases allow the Paperless Office administrators to change the system without modifying the scripts.

3.2 Script Structure

Furthermore, all scripts generate HTML dynamically. This is to say that all the scripts are written in PERL, and any HTML that is generated is produced from the PERL. The only static HTML pages are those that contain help information.The scripts are also written to be recursive. All functionality for the script lies in one file, which makes editing and management much easier.

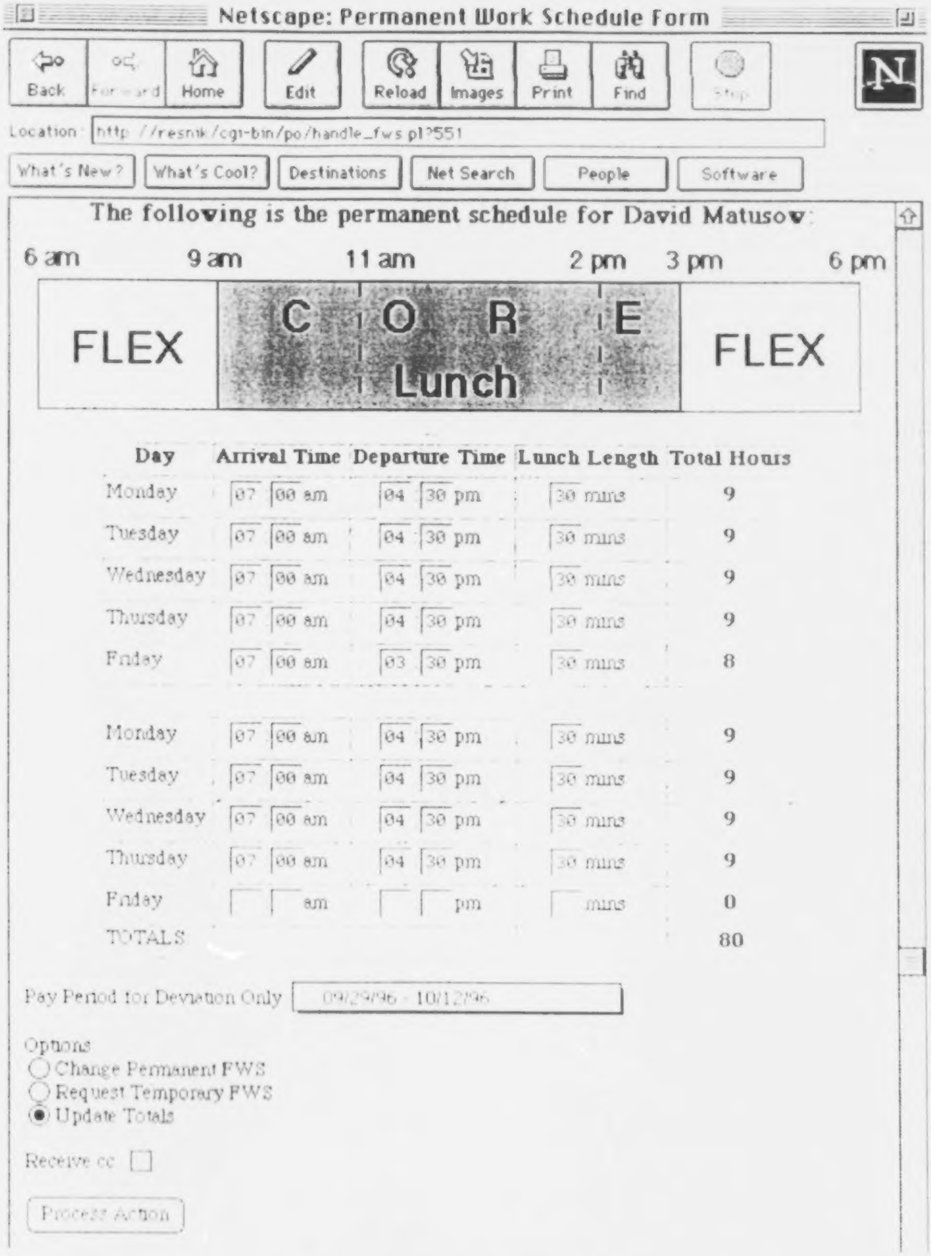
3.3 The Leave Slip

In the government, when you are not going to report for work, due to sickness, vacation, or some other reason, an employee is required to submit a leave slip to his/her supervisor for approval. Traditionally, this was accomplished by a paper form which was filled out and handed to the supervisor who would sign it, and forward it to the timekeeper for processing. Through the Paperless Office, the leave slip was the first of the on-line forms to become operational. Now, an HTML form (see diagram 1) is generated and the user fills out the form with some validation checks built in. When the user is ready to submit the leave slip, an Email is automatically generated and sent to the proper supervisor for that employee.

Diagram 1. The leave slip form in HTML

3.4 The Flexible Work Scheduler

Most businesses around the globe are realizing that society wants to spend more time at home with their families. NASA is no exception, and recently introduced a Flexible Work Schedule (FWS) plan. How to manage this administrative function then was passed down to the individual areas. In the FDD, we decided to manage this function through the Paperless Office. Under this system, management can determine when employees are working, and when they are off. The employee logs into the HTML form (see diagram 2) and enters their schedule. This data is then stored in a database, and an Email is generated when any change is requested. This Email gets routed to the employee's supervisor for approval and notification. They are easily able to manage employee's schedules because all the information is online, and easily searchable. Finally, the system allows individual employees to query the database to find out who is off on which days. This helps project managers in scheduling and planning team meetings.


Netscape: Permanent Work Schedule Form
Back
Forward
Home
Edit
Reload
Images
Print
Find
Stop
N
What's New?
What's Cool?
Destinations
Net Search
People
Software
6 am
9 am
11 am
2 pm
3 pm
6 pm
FLEX
C O R E
Lunch
FLEX

Location: http://resnik/cgi-bin/po/handle_fws.pl?551

The following is the permanent schedule for David Matusov:

Day	Arrival Time	Departure Time	Lunch Length	Total Hours
Monday	07 00 am	04 30 pm	30 mins	9
Tuesday	07 00 am	04 30 pm	30 mins	9
Wednesday	07 00 am	04 30 pm	30 mins	9
Thursday	07 00 am	04 30 pm	30 mins	9
Friday	07 00 am	03 30 pm	30 mins	8
Monday	07 00 am	04 30 pm	30 mins	9
Tuesday	07 00 am	04 30 pm	30 mins	9
Wednesday	07 00 am	04 30 pm	30 mins	9
Thursday	07 00 am	04 30 pm	30 mins	9
Friday	am	pm	mins	0
TOTALS				80

Pay Period for Deviation Only: 09/29/96 - 10/12/96

Options
○ Change Permanent FWS
○ Request Temporary FWS
● Update Totals

Receive cc ☐

Process Action

Diagram 2. The FWS form in HTML

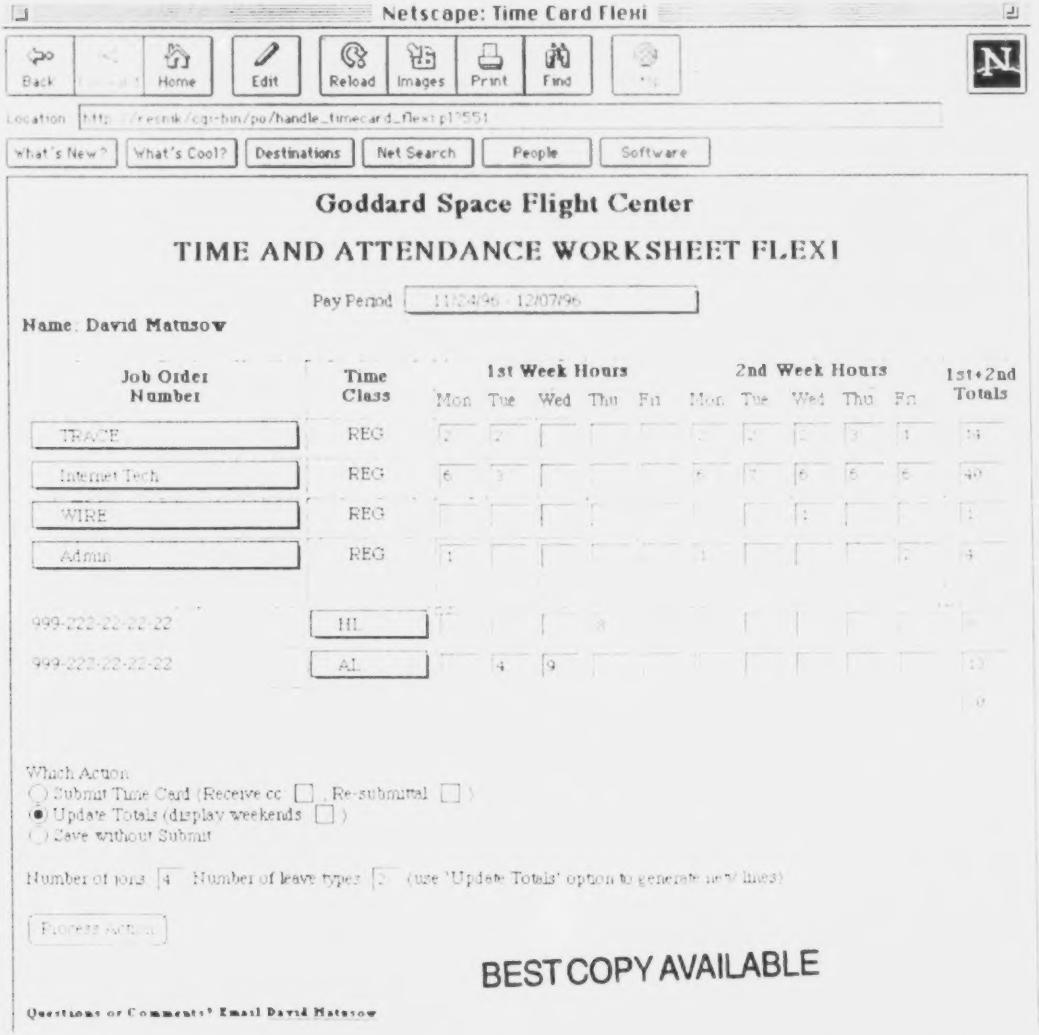
3.5 The Time and Attendance System

One of the largest areas where administrative functions cost the most and are the hardest to control is in the area of handling employees' timecards. In the Government, personnel are required to submit bi-weekly timecards to the payroll office. In most areas, this is accomplished in a very tedious manner. First an employee fills out a paper timecard and hands it to his/her supervisor who validates and approves the timecard. Then, the timecard is filled out by the

BEST COPY AVAILABLE

timekeeper, who gives the card back to the supervisor for signature. The process is tedious and requires many physical interactions, and it is easy to see why this costs so much money.

In the FDD, we have moved as much of this process as is currently permitted to the Paperless Office. We now have a flexible timecard system whereby the user logs in and enters their time and attendance information in an on-line HTML generated form (see diagram 3 below). The user has the option of saving the information, or submitting it for processing. When a timecard is submitted, an Email with a representation of the timecard gets sent to the timekeeper for processing. This new system is very user friendly, allows for a timecard system that anyone can access because it is platform independent, and automates much of the process saving time and money. Other benefits include being able to use the database to gain statistical data on what projects employees are charging. This function is a universal requirement on management, and frequently a time consuming one. In the past, a business would have to go through all the paper forms and enter the data into a database for analysis. Now, it is a simple matter of selecting the information, and the web browser automatically displays the appropriate data.



Goddard Space Flight Center

TIME AND ATTENDANCE WORKSHEET FLEXI

Pay Period 11/24/96 - 12/07/96

Name: **David Matusow**

Job Order Number	Time Class	1st Week Hours					2nd Week Hours					1st+2nd Totals
		Mon	Tue	Wed	Thu	Fri	Mon	Tue	Wed	Thu	Fri	
TRACE	REG	2	2				2	2	2	3	1	14
Internet Tech	REG	6	3				6	7	6	6	6	40
WIRE	REG								1			1
Admin	REG	1					1				2	4
999-222-22-22-22	HL				8							8
999-222-22-22-22	AL		4	9								13
												0

Which Action
○ Submit Time Card (Receive cc ☐, Re-submittal ☐)
● Update Totals (display weekends ☐)
○ Save without Submit

Number of jobs 4 Number of leave types 2 (use 'Update Totals' option to generate new lines)

Process Action

Questions or Comments? Email David Matusow

BEST COPY AVAILABLE

Diagram 3. The Timecard form in HTML

4.0 Conclusions

Many organizations spend vast amounts of time and money searching for ways of becoming more steam-lined and efficient. In the FDD, we have seen that a great deal of savings can be realized thru the use of the Internet. One of the greatest aspects of the Internet is its cross platform nature. It doesn't matter whether you have a workstation or a personal computer, or whether you work in an office or at home, if you have an Internet connection and a web browser, you need no other special software. Often times when companies try to save money, they procure COTS software to provide a solution to their problem, but usually this creates more problems than it solves. It will only run on certain platforms, and it doesn't necessarily provide the same capabilities on all platforms, and it may not be user friendly. Through the Paperless Office system, we have found that for a relatively small cost (about 1 staff month of effort), we can create the functionality to handle the routine administrative functions that can usually bog down an organization. In the FDD alone, we have seen processing time drop from 10 hours to 2 hours on the time and attendance process.

Furthermore, we have found that we can create customized, flexible, and efficient applications thru the use of PERL and HTML. CGI scripting has given us the ability to create dynamic web applications that can handle almost all needs required by these administrative functions. Also, we have noticed that thru proper software engineering techniques, an organization can obtain a great deal of reuse of these scripts, further saving time and money.

Many organizations today are creating their own Intranet as an internal information exchange system. The FDD's experiences show that companies can move beyond an information bulletin board and utilize an Intranet as an interactive administrative tool as well.

The bottom line is that everyone knows about the Internet, and having an Intranet, but the question is: **"are you maximizing your Intranet?"** Through mechanisms like the Paperless Office, cost reduction and efficiency can be realized, with the additional benefits mentioned above.

On Creating Hierarchical, Interlinked FAQ Archives

Tom Murphy, Silicon Graphics, USA, murphy@sgi.com

Abstract: An archive of frequently asked questions (FAQs) can be a lifesaver for those seeking knowledge. FAQs initially decreased the signal-to-noise ratio of newsgroups. They are now used in any situation to help a person quickly gain the knowledge and experience of others. The information maintainer has the challenge of providing the most current information with the least amount of maintenance effort. A fundamental trade-off exists between ease of maintenance and ease of use. Significant increases in usefullness typically require significant levels of additional maintenance work. This paper discusses ways to automate based on use of a FAQ compiler.

Introduction

In this paper, we will explore some of the desired characteristics in the creation and use of FAQ archives, as well as briefly examine the solutions of other individuals. We will examine a particular solution involving the design goals and implementation of a FAQ compiler. Finally, we will consider planned and desired extensions to the FAQ compiler.

Elements of Good FAQ Creation and Use

When creating a FAQ compiler, you should try to incorporate the following attributes:

- Educational utility:
  - Presentation formats:
    - HTML for ease in viewing and transmission efficiency
    - ASCII for fastest transmission, information excerpting, full text searching and accessibility
    - Neutral FAQ format used for specific presentation formats, such as the above two

  - Information access:
    - Possibility for multiple hierarchical presentations of the individual questions
    - Ability to see changes/additions from the broadest perspective, that is, the top FAQ, down through the hierarchy to a question, down to the changed or added word, as desired by the user
    - Changes relative to the last time the user viewed the changed item
    - Extensive cross referencing so the user can most easily find the information sought
    - FAQWA — frequently asked questions without answers, which anyone can augment
    - Anyone can take existing data and polish it into a FAQ question and answer
    - Supply with each question: dates of creation, expected update and expiration, name of author, home page and email address

  - Content:
    - A system that allows anyone to add or change data (no one person is the bottleneck)
    - All the USENET FAQs
    - All FAQs having uniform access methods and features independent of a particular author

- Ease of mechanism:
  - Ease of maintenance:
    - Easy to produce FAQ
    - Ability to remove questions automatically if not updated
    - Ability to remind authors automatically to update their questions
    - Ability to tailor question removal and author reminders to characteristics of a particular question
    - Author errors automatically routed to author rather than just the FAQ maintainer
  - Ease for author/editor:
    - Information defined in only one place
    - Possible for anyone to change data while viewing the FAQ. Author alerted of all changes
    - HTML used to structure question
    - Easy to insert new questions
    - Easy to shuffle existing questions within and between FAQs
    - Easy to do cross references

Other FAQ Solutions

The canonical FAQs, which are available by FTP from rtfm.mit.edu [MIT 97], are basically ASCII FAQs typically posted monthly to USENET newsgroups. While some FAQs post a list of changes, most leave it to the reader to determine what is different since the last posting. Whereas some FAQs add new questions to the end to simplify noting additions, others integrate new questions within appropriate places within the FAQ.

Both Ohio State [Ohio 97] and the Oxford University Libraries Automation Service [Oxford 97] have added some HTML to each FAQ and one or more TOP level hierarchical entry points to the FAQ archive. Both use WAIS [Pfeifer 95] search engines; both use multilevel table of contents hierarchies for progressive disclosure of available FAQs. For instance, one level for each component in the name of the newsgroup. Both organizations, at times, divide the FAQ into the parts posted to the newsgroup or transferred by FTP from rtfm.mit.edu. Both treat FAQs as a monolithic formatted datum, that is, a <pre> at the front and a </pre> at the end. Both of them ferret out URLs in the body of the answer and anchor them to links. For example, consider the fixed format ASCII and the corresponding HTML (at http://www.cis.ohio-state.edu/hypertext/faq/usenet/www/faq/faq.html and http://www.boutell.com/faq/ [Boutell 96]) versions of the World Wide Web FAQ.

Ohio State also supports two FAQ formats that yield a much more HTML-like product. One exploits RFC 1153 [Wancho 90] where each question and answer is "digested" as a separate post. This allows each question to be a separate document, which is a significant improvement.. The WAIS search is just of newsgroup names, archival names, subjects, and keywords. Ohio State has two versions of the FAQs, one organized by alphabetical by title of FAQ and the other by newsgroup. The alphabetical one has a page listing links to all the questions of a particular FAQ. A particular question has links automatically placed at the bottom to the preceding and next question as well as to the table of contents of this particular FAQ.

The Oxford University Libraries Automation Service style of automation is to form two different table of contents into the FAQs, one by FAQ name and the other grouped topically. They cleverly pick up the USENET newsgroup purpose description as a title to the FAQ.

Setext [Feldman 97] offers an interesting twist on maintaining both ASCII and HTML. It relies on unobtrusive natural structuring of ASCII, which is then mapped to LaTex [Lamport 86], by latex2html [Drakos 96].

FAQ Compiler Design Goals

You should try to incorporate the following attributes into the design of your FAQ compiler:

- *Ability to allow anyone to add, delete, and modify questions*. This helps to maintain currency. To support this well, I had to maintain a history of changes, so I could back out someone's "help" if necessary. The SGI source code maintenance system helped here—you change it, you own it.

- *Ability to support a short developer Makefile*. This eases peoples deep seated hatred of Makefiles.

- *Ability to generate HTML or ASCII for particular source file*

- *Fast compilation*

- *Define each datum in only one place*. For instance, question titles and FAQ titles are referred to in multiple places—in questions, in the FAQ table of contents (TOCs), and in the questions and TOCs of other FAQs. Yet, a title appears in only one place—in questions, it's the first line in the question file; in FAQs, it's the first line in a TOC file. If data is defined in only one place, then one version cannot be out of sequence with the other.

- ***Ability to maintain both HTML and ASCII.*** ASCII has its fundamental speed and searching advantages over the load time of an HTML viewer and the desired documents. It is also much more accessible for people with disabilities [W3C 97].

- *Ability to maintain multiple versions of a FAQ*. Users should be able to tailor the view of a FAQ. Examples include showing the recent changes (additions in green, removals in strikeout). You can also choose to put all questions and answers in one or several documents.

- *Ability to support a hierarchical FAQ structure*

Implementation

The FAQ compiler operates on two kinds of source: FAQs expressed in internal format and ASCII FAQs.

Source Processing

The overall goal is to define data in one spot, which helps to simplify the ability to keep questions and answers current and accurate. All of the extensions look similar to HTML. In fact, keywords of the form <<FileName>> are similar to the HTML keyword <p>; macros of the form &&q; look a lot like the HTML macro &amp;. The intent is that the extensions "feel" as much like HTML as possible.

The <<File Name>> construct compiles to an anchor and its anchored text—that is, a URL reference. Text may be supplied within the anchor to define both the text for the HTML form and the ASCII form. If no text is supplied, you get the question in the HTML case and the question number in the ASCII case; the URL is printed in parentheses in the ASCII case. A variant of this construct that does not generate an anchor exists, but it just allows differing text between the HTML and ASCII versions.

The &&q; construct compiles to text describing one of three attributes of a question:

- t, the title (the question itself)
- q, the question number
- c, the color of the question

Possible question colors are green (new question), yellow (changed question), and black (unchanged question). The same colors are used to color a FAQ: green (the entire FAQ is new), yellow (there has been a change in the FAQ or sub-FAQ), and black (no change anywhere in the FAQ.) You can set the change period to be anything. For instance, I set it to one month. The color text is that of an <img> reference to an image of a colored ball. There is one other form of this construct, &&h;, which expands to the URL of the anchor in which it appears. If no extensions appear within an anchor, then the URL is placed in parentheses at the end of the anchor (applies to ASCII version only).

The compiled FAQ hierarchy, seen through navigation links between FAQs, is inherited from the FAQ directory hierarchy of the source tree. FAQ directory names are constrained to be unique, thus compiled FAQs can be stored in a single directory, maintaining FAQ URL integrity when FAQs are moved around in the source tree. The file name of a question is obviously unique within a directory. It is used as the name of the question. Thus it takes only a FAQ name and question name to uniquely reference any question in the FAQ source tree.

The table of contents file, TOC, is processed specially. The <<File Name>> construct refers to questions in this FAQ and possibly to other FAQs. The TOC file expands to an image reference to a colored ball reflecting the age of the question or FAQ; TOC also expands to the title of the question or FAQ. A syntax extension that supports an external FAQ is already in HTML form—it gets a different <img> and compiles to an anchor referencing the FAQ. ASCII FAQs also get their own FAQ icon.

The FAQ compiler also handles ASCII FAQs in a simplified way. The compiler accepts a pattern to define either the divider between FAQs or a pattern to match the beginning of a particular FAQ, and it presumes a table of contents before the first question. The compiler can process FAQs directly from the Internet.

Four Lex/YACC [Lesk 75] [Johnson 75] grammars form the core of the FAQ compiler. One processes the TOC file into canonical FAQ compiler source, and it uses the TOC to define identity and sequence of questions that make up this particular FAQ. The second grammar, which is the real workhorse, understands the complexity of both the extended HTML and the relevant parts of the HTML grammar. The second grammar maps the file to a raw neutral form with one word per line. This raw file is compared with its partner from the past month to define the changes. Finally, the program maps the raw neutral file, based on the comparison, to form the neutral output file. The next two grammars each take a neutral output file and map it to a specific output format: HTML or ASCII. The HTML format is lightweight (since the neutral output format is already similar to HTML); the ASCII format is quite complicated because it has to understand much of the HTML grammar.

Makefile

The standard Makefile in the FAQ source tree looks like this:

```
#!smake
include $(ROOT)/usr/include/make/faqcommonrules
```

The simplicity of this file appeals to those who tend to dislike Makefiles. faqcommonrules rewrites a special Makedepend file whenever the TOC file has been changed. faqcommonrules then re-invokes itself with the special Makedepend file to decide which questions need to be translated to internal form, including a comparison with last month's internal form, if it exists. Finally all the internal forms are processed to form the HTML FAQs and then again to form the equivalent ASCII FAQs. Many Makefile intricacies within the Makefile include file faqcommonrules exist, as well as a number of helper programs to make the basic Makefile look so simple.

ASCII Viewer

To keep with the lightweight nature of ASCII, ASCII viewers should provide fast execution. The viewers should also have a robust interface; for instance, Q43, q43 and 43 should all be accepted as names for question 43. The FAQ viewer tries to emulate as many of the capabilities of an HTML viewer as possible, while keeping to its lightweight nature. The FAQ compiler produces a sorted command file for each FAQ directory. It defines all the standard commands, as well as four names for each question: the question number, the question number preceded with a 'q' or 'Q' and the file name of the question, which appears to the user to be a sort mnemonic for the question. Thus the FAQ viewer need only do a log(n) search for each command. The FAQ viewer just needs to identify the correct file name, which is the second entry on each FAQ compiler written command file. The file is copied to a workstation (faster than NFS mounting) and passed to a pager.

Future Enhancements

My initial FAQ solution preceded the existence of the World Wide Web. Its current second generation uses the power of the early Web (HTML 1.0.) It is well in need of its third generation to exploit the power of CSS [Lie & Bos 96] and other features of the current Web.

Here's a list of possible enhancements:

- Latest HTML extensions added to ASCII generation, most notably tables.
- Use of CSS to replace the HTML extensions. This will allow any HTML editor to operate on FAQ source. XML [Bray & Derose 97] will ultimately provide the most clear method for expressing the extensions.
- Use of NIF-T-NAV [Jones 96]. It will allow a theoretically larger table of contents which will initially be quite brief. This allows fewer levels in the hierarchy and thus more ease in navigation for the user. I will also provide a traditional static hierarchy, which works better with using the viewer for searching in the current page.
- Update source from HTML rather than the Silicon Graphics source code maintenance system. While the company's source code system is typically easy for engineers, it tends to be a barrier to those who don't know it.
- Various topical TOCs to provide different "windows" into FAQ data. This is a time-intensive project ripe for a process to make it achievable.
- Identification of identical questions in Internet FAQs from month to month so they can be completely presented as FAQs in FAQ compiler format; also, identification of embedded URLs in Internet FAQs
- Automatically annotate each question with its last modification date, the author's name and e-mail address (automatically linking to an appropriate home page, if one exists), and expected next modification or expiration date
- Hierarchical WAIS search. Right now, WAIS searches all FAQs. There are two additional search paradigms to add. One is the option to search the current FAQ along with any subFAQs. The other is to use NIF-T-NAV to organize the selection of the FAQs that should be searched (requires a search-oriented version of NIF-T-NAV).
- Instead of eight versions of each question, use a cgi-bin script to weave a particular desired view from a single neutral display file

- Use of existing HTML to ASCII translator. It is cumbersome to maintain a YACC/LEX grammar for the HTML moving target. It also problematic to use someone else's product which may also lag and which you may have to do some customizing to use. I recommend problematic over cumbersome in this particular case.

Conclusion

The FAQ compiler was born out of the caldron of necessity. It was not designed to be the general answer to automatic FAQ generation. However, it has proved to be a useful answer to many of the questions that arise in our high-technology environment. Much room exists for more powerful FAQ generation tools to meet what I find desirable in a FAQ system; even more room for tools we together find desirable in an FAQ system.

References

[Boutell 96] Thomas Boutell, *World Wide Web FAQ*, http://www.boutell.com/faq/

[Bray & DeRose 97] Tim Bray and Steve DeRose, *Extensible Markup Language (XML): Part 2. Linking* (working document), World Wide Web Consortium, http://www.w3.org/TR/WD-xml-link

[Drakos 96] Nikos Drakos (1996), *latex2html, LaTeX to HTML translator*, http://cbl.leeds.ac.uk/nikos/tex2html/latex2html.tar-gz

[Feldman 92] Ian Feldman, *Setext, a mark-up language*, http://www.bsdi.com/setext/setext_concepts_Aug92.etx

[Johnson 75] S. C. Johnson, *Yacc - Yet Another Compiler-Compiler*. Comp. Sci. Tech Rep. No. 32, Bell Laboratories, July 1975

[Jones 96], Kirsten L. Jones, *NIF-T-NAV: A Hierarchical Navigator for WWW Pages*, Fifth International World Wide Web Conference, http://www5conf.inria.fr/fich_html/papers/P39/Overview.html

[Lamport 86] L. Lamport (1986), *LaTeX - A Document Preparation System*, Addison-Wesley

[Lesk 75] M. E. Lesk, *Lex - A Lexical Analyzer Generator*, Comp. Sci. Tech. Rep. No. 39, Bell Laboratories, October 1975

[Lie & Bos 96] Lie, Håkon Wium and Bos, Bert (1996), *Cascading Style Sheets, level 1*, World Wide Web Consortium, http://www.w3.org/TR/REC-CSS1-961217

[MIT 97] Massachusetts Institute of Technology, *rtfm.mit.edu USENET FAQs*, ftp://rtfm.mit.edu/pub/usenet/

[Ohio 97] Ohio State, *USENET FAQs*, http://www.cis.ohio-state.edu/hypertext/faq/usenet/

[Oxford 97] Oxford University, *Oxford University Libraries Automation Service World Wide Web Server*, http://www.lib.ox.ac.uk/internet/news/

[Pfeifer 95] Ulrich Pfeifer (1995) *freeWAIS-sf*, ftp://ftp.germany.eu.net/pub/infosystems/wais/Unido-LS6/freeWAIS-sf-2.0/fwsf.ps.gz

[ratcliffe & Murphy 97a], ratcliffe, dave and Tom Murphy, *Developer Toolbox*, https://www.sgi.com/toolbox//

[ratcliffe & Murphy 97b], ratcliffe, dave and Tom Murphy, *Taste of the Toolbox*, http://www.sgi.com/TasteOfDT/

[Wancho 90] F. Wancho, *RFC 1153, Digest Message Format*, http://www.cis.ohio-state.edu/htbin/rfc/rfc1153/

[W3C 97] World Wide Web Consorium, *Web Accessibility Initiative*, http://www.w3.org/WAI/

Acknowledgments

My partner, dave ratcliffe, helps my job be quite enjoyable, for work and friendship are most nicely intertwined. We take a lot of pride in the SGI Developer Toolbox [ratcliffe & Murphy 97a] and the Taste of the Toolbox [ratcliffe & Murphy 97b], which we produce.

Editorial assistance provided by Anatole Gordon.

High Performance Web Server System: TCP Based Switching Approach

B. Nandy, K.L.E. Law, A.S.J. Chapman

Nortel Technology
PO Box 3511, Station C
Ottawa, Ontario K1Y 4H7
Canada
email: {bnandy, eddiel, achapman}@nortel.ca

Abstract:

This paper addresses the performance issues of a web server. A network architecture of servers is proposed for high throughput and fast response time. The client requests are distributed among multiple servers in a transparent manner by an intermediary device. The distribution of requests are performed at the granularity level of TCP sessions allowing the load balancing among servers. A preliminary software implementation shows that the proposed approach can improve the throughput as well as the response time.

1.0 Introduction

Last few years have observed a phenomenal growth in Web usage. This growth has caused a significant research interest in improving the performance for Web Systems. At a macro level, Web system consists of three components: Client, Communication Protocol [1.]and Server. Efforts are being made at each component to enhance the performance of the overall web systems. In this work, we describe a multiple server architecture which can meet the increased demand of web traffic.

A powerful web server may be developed by improvement of different components of a web server (e.g., CPU speed, disk performance, file system performance, performance of TCP/IP, server software architecture etc.). Alternatively, multiple servers can be used to handle high rate of server requests. Different possible approaches to multiple server systems are briefly outlined:

The use of Domain Name Server (DNS) to distribute requests among multiple servers was done at NCSA[2.]. This approach resolves the logical IP address and maps to one of several physical addresses in a round robin manner. It has provided some success in distributing the server load. However, the approach could not balance the load among servers. Another problem with this approach is, once the IP address resolution is cached in the local memory, the client may never contact DNS.

Another possible approach is to use the HTTP[3.] level redirection capability to move requests among multiple servers. However, this requires a round trip delay between the client and server before the request is redirected to a different server. Moreover, if the first server is already very busy, the response delay will be even greater.

Other possible approach is to use an intermediate router-like device which distributes IP datagrams or TCP segments among multiple servers (the unit of transfer for TCP between machines is called Segment). A mechanism (a possible hashing function) can direct traffic to different servers based on the IP addresses. Alternatively, TCP session also can be chosen as the unit of switching. In this work, TCP segments are distributed among multiple servers using an intermediate device.

Section 2 describes the TCP based server switching approach. Section 3 discusses different issues associated with TCP based switching. Section 4 presents the implementation results. Section 5 discusses the usefulness of the proposed approach for Web application and the future work.

2.0 TCP based server switching approach

A HTTP session is aggregation of one or more TCP sessions. A router-like system called a "depot" sits transparently between the clients and the servers (Figure 1 shows a typical depot deployment scenario). The depot forwards the TCP sessions among multiple servers based on the server load balancing criteria. A TCP session consists of multiple TCP segments. All the TCP segments of a given TCP session are served by the same server.

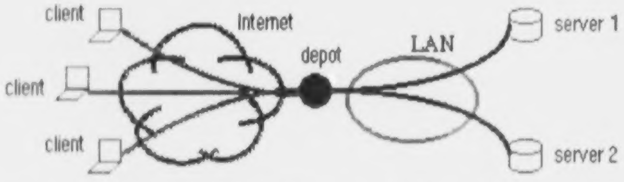


Figure 1: A Typical Client-Server Configuration with Depot

HTTP is a stateless protocol. A web server obtains everything it needs to know about a request from the request itself. After the request is serviced, the server can forget the previous transaction. Thus, each request in HTTP is disjoint. If all the servers are identical (or sees the same file system using a distributed file system), the server from which the request is served is of little relevance to the end user. Different TCP sessions can be allocated to different servers without knowing if all the TCP sessions belong to same HTTP session or not. Thus, TCP based server switching allows a nice granularity for load balancing among multiple web servers.

2.1 Server Switching Architecture

The depot is a forwarding device between the clients and the servers. All the clients access the server system using the depot IP address. The depot does not generate any TCP segments. The distribution of TCP sessions among multiple servers by the depot remains transparent to client. Since, the forwarding of all TCP segments of a given TCP session must go to the same server, a mapping between client IP and port with server IP and port is maintained. The entry for this mapping is to be maintained in the depot by following the TCP protocol. This entry is preserved in the depot as long as there is a possibility of arrival of a TCP segment from a client or a server. Depot has the following functions: (i) inspect all segments in both directions at IP and TCP levels, (ii) choose a server based on load balancing criteria for a new TCP session, (iii) forward TCP segments for existing sessions to the already chosen server, (iv) forward TCP segments from servers to the clients, (v) clean up the mapping entry when TCP sessions end and (vi) watch for and handle anomalous TCP segments.

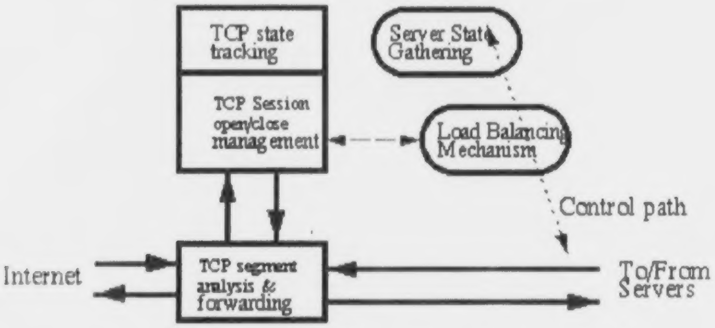


Figure 2: The Depot Functions

For a new session the TCP segment analyser identifies the TCP connection setup request and forwards an information to the session management block. This block is preloaded with the next server to be allocated. The choice of server may be anything from a simple round robin to a complex load balancing algorithm on the basis of knowledge of the server statistics and network states. Server probing can be performed periodically to obtain the server statistics. The TCP segments carrying data are forwarded by the TCP segment forwarding block by identifying the entry of mapping list between the server and the client. The tracking of TCP states for both the clients and the servers are performed to facilitate the management of TCP session closing.

The depot maintains a mapping list of all the active TCP connections in a table called the primary table. Each connection between client and depot is identified by the combination of client ip address and port number. All the connections from the client to the depot comes to a predefined port. Each connection between the depot and servers is identified by the port number assigned at the depot and the server IP address. All the connections to servers come to the same predefined port. The incoming segments from different clients may contain same source port number. Thus, new port numbers are assigned by the depot before forwarding the segment to the server. A typical primary table entry has the following: client ip address and port number, server IP address and assigned port number at the depot, TCP states and related parameters (ack number, seq number etc.).

The depot maintains another list of connections in a table called secondary table. The entries from primary table are moved after close of connections under normal or anomalous conditions. The secondary table entries are maintained for 2MSL (i.e., 2 minutes for this implementation) period. The 2MSL is important since a TCP segment may arrive after a connection is closed due to variable delay at the network.

The reason for maintaining two separate tables at depot is to reduce the search space for finding a match on arrival of a TCP segment. The number of entries in secondary table is very large since the entries are maintained for 2MSL period. The entries in primary table are the only active TCP connections. Most of the incoming TCP segments should, should find a match in the primary table.

TCP segments flowing in both the directions are analysed in order to get a complete view of the state of the session. The depot guesses the states in client and server. The guessed states at depot may be different from the actual states at the TCP termination points, if a TCP segment is dropped or corrupted in the network between depot and the client or server. The state tracking is necessary to manage primary table entries on receipt of a reset (RST) segment or closing of a TCP session.

2.2 Handling of different TCP segments

Arrival of different TCP segments are handled in the following manner:

- SYN: A SYN segment (client ip address and port for client segment, server ip address and depot port for server segment) is matched with the entries in primary table and then with the secondary table. If no match is found (i.e., arrival of a new connection), a new entry is created in the primary table and based on the load balancing criterion, a server is allocated for the connection. If a match is found (i.e., duplicate SYN), the segment is forwarded to already allocated server/client.

- FIN, PSH, URG: The match is found from the primary or secondary table and forwarded to the appropriate server/client. If no match is found, the segment is dropped.

- ACK: All ACK segments are forwarded to the server/client if an entry is found in the table. If the ACK segment causes the state transition to TIME_WAIT state, the entry is moved from the primary table to secondary table for 2MSL time-out.

- RST: All reset segments are validated by checking if the sequence number is in the window. If the state is SYN-SENT, the RST is considered valid if the ACK field acknowledges the SYN. If the RST is valid, the entry from the primary table is moved to secondary table.

The exception conditions are handled in the following manner:

- If an entry in primary table is inactive for a long time (say, more than 20 mins), the entry is moved to secondary table. This is necessary, since client or server may crash without proper termination of a TCP session and cause an entry to the primary table to remain for ever.

- If an entry in tables (primary and secondary) is not found on arrival of a segment other than SYN, the segment is dropped. This is necessary, since, depot does not know the destination for the segment.

- Depot does not cause any interruption of the ongoing TCP session between the client and the server. This is because, depot forwards every segment irrespective of the guessed states at depot as long as an table entry is found. For example, an RST may be lost in the network between depot and client but the entry is still maintained in secondary table which enables the forwarding of any retransmitted segment.

3.0 Different Issues Associated with TCP Based Switching:

State Information Issue:

The current thrust to introduce state information in the HTTP transaction [4.][5.] requires some analysis of the proposed approaches. Both the ideas of HTTP State-Info mechanism by Kristol[4.] and Cookies from Netscape[5.] work with the concept of saving the server state in the client. A server, when returning an http object to a client, may also send a piece of state information which the client will store. Included in the state object is a description of the range of URLs for which that state is valid. Any future HTTP requests made by the client which fall in the range will include the current value of the state from the client back to server. In netscape the state object is called Cookie.

This state information adds power which enables a new type of web application. For example, when one browses through a "virtual shopping mall" and add items to buy from a list of items, and pay for all the chosen items at the end will require the state-information for the chosen items. The model of TCP connection based server switching will work well even with the new concept of state-information in HTTP since all the states are stored in client and HTTP protocol is still essentially stateless.

Authentication Issue:

HTTP provides a simple challenge-response authentication mechanism. A server which requires clients to authenticate themselves replies to a client's HTTP request with an Unauthorized (401) error code with indication of the required authentication method. The client may respond with an authentication information called "credentials". The domain over which the credentials can be applied is determined by the "protection space".

In a typical implementation, encountering an Unauthorized (401) response causes a browser to prompt the user for authentication credentials. For any subsequent access to the same protection space, the browser may cache the user credentials and automatically include them in its access of server. This is called Basic Authentication Scheme. The password is matched at the server based on the access control lists to the stored password file for the user.

In TCP based switching, different file requests from the same HTTP session will get forwarded to different servers. Thus, identical access control lists and password files are to be maintained in all the servers. Alternatively, a distributed file system (like AFS[8.]) can be shared by all the servers to access the password file and access control list.

Secure Web Server Issue:

Secured Socket Layer[6.] protocol is a crypto enhanced version of TCP/IP, developed by Netscape Corporation. The problem may occur, with the introduction of SSL protocol on top of TCP layer. An SSL session is stateful, the SSL handshake protocol coordinates the state of the client and the servers. The complete SSL handshaking session is on top of a single TCP connection. Thus SSL handshaking protocol works perfectly and the keys get exchanged between the client and server. But the key remains with one server. Thus, any subsequent TCP session from the same client if forwarded to a different server is impossible to decrypt. A common server is necessary where the keys can be saved and forwarded on request from a server.

4.0 Experimental Results

A pentium PC running NetBSD operating system is used for the software implementation of Depot. The clients and servers are connected to depot using 10Mbps Ethernet. The client requests are generated using a benchmark software from Zeus corporation [7.]. The traffic is forwarded by depot to two identical NCSA/1.5.1 Web servers in a round robin fashion.

Server throughput (total bytes transferred per second) is measured for each test case. The total number of bytes of data and the http headers divided by the time taken to transfer indicates the server throughput. The number of requests served by the server per second is also measured. Both the measures include a variable network delay. To minimise the variable network delay, the experiment is performed at a time when the network is very lightly loaded.

The experiment is performed by retrieving files of three different sizes: 100 bytes, 1 Kbytes and 10 Kbytes. The number of simultaneous client requests are fixed at different values between 1 to 100. The number of concurrent connections for a given test is always maintained at a fixed number. As soon as a connection is closed (normally or abnormally), the client software initiates another connection. The total number of client requests for each test is 1000. The experiment is performed with a single server and a depot system with two servers.

Table 1: Server Throughput in Kilobytes per second

concurrent connec-

tions

file size: 100 bytes file size: 1 Kbytes file size: 10 Kbytes single server depot system single server depot system single server depot system 118.2217.4666.5856.6752.1351.98

529.2144.6299.2680.75151.85168.18 1021.155.1394.8399.81226.80244.53
1510.1839.565.75101.73223.1277.06 2011.1729.8642.64111.93251.24289.91
256.4630.0033.86117.53233.73297.89 1006.5512.23X37.59X282.81

Table 2: Served Requests per second

concurrent connec- tions	file size: 100 bytes		file size: 1 Kbytes		file size: 10 Kbytes	
	single server	**depot system**	**single server**	**depot system**	**single server**	**depot system**
1	64.61	61.92	55.16	46.95	5.00	4.99
5	103.48	158.23	82.19	66.88	14.55	16.13
10	74.59	194.70	78.31	82.55	21.68	23.44
15	35.73	138.95	54.03	84.21	21.35	26.53
20	39.03	104.62	34.87	92.29	23.97	27.67
25	22.48	105.22	27.55	97.23	22.40	28.49
100	22.41	40.79	X	28.51	X	25.69

1. For one client request (any file size) at a time, depot system is always slower than a single server. This is due to the store and forward delay introduced at depot (The store and forward delay incurred at depot is 0.65 mili-second). 2. For multiple simultaneous requests, the requests are served in parallel by two servers. In general, depot system shows an improved throughput and served requests per second. Another reason for performance improvement is the reduced workload at each server due to the distribution of requests. 3. For 100 bytes and 1 Kbytes files, the single server performance degrades quickly with the increased number of simultaneous connections. The performance with depot system is consistently better than the single server system. For 10Kbytes file, the single server performance remains flat. However, the performance with depot system is better than a single server. 4. For 100 simultaneous requests, the single server could complete the test for only 100 bytes file size. The depot system could complete the tests for all three file sizes. 5. It is found that the throughput and requests served per second with depot system are more than double than a single server system with large number of simultaneous connections. For example, the number of requests served per sec for the test case of 1 Kbytes file size with 20 simultaneous connection is 92.29. This is 2.65 times improvement over a single server. This super-linear improvement is due to the halved number of simultaneous connections (i.e., 10) on each server. The number of requests served per second by a single server with 10 simultaneous requests is 78.31. Thus, a maximum served requests of (2*78.31) 156.62 is theoretically attainable. 6. A decrease in the number of served requests per second with

the increased file sizes were observed. The possible reasons are (i) as the file size increases, the time taken to complete a request also increases, (ii) other bottlenecks(e.g., disk I/O limit, transfer rate at ethernet interface) start affecting the result.

5.0 Discussion

The appeal of universal connectivity and ease of access to information are the major factors for the phenomenal growth of web applications. Another reason for growth is, the business community has accepted web as the medium of communication of their products and services to customers. The quick growth along with the application like electronic commerce has necessitated a web service with good response time, high availability and security. The proposed web server system is one component of the complete web application towards the above mentioned requirements. It has been shown that the proposed solution can handle higher volume of traffic with an improved response time.

The approach of TCP based switching is also applicable to the idea of persistence sessions with the server. This will cause a coarser granularity of load distribution among servers (e.g., instead of the flexibility of switching among multiple short tcp connections per http page, a single long connection will persist). However, even if the states are saved at server side, there is no issue as long as the states are used during the same TCP connection. At the same time, the length of tables (primary and secondary) will be shorter due to a lesser number of TCP connections.

6.0 References

1. D.E. Comer, Internetworking with TCP/IP, Volume 1: Principles, Protocols, and Architecture, 3rd Ed., Prentice Hall 1995. 2. E.D. Katz, M. Butler, R. McGrath, "A Scalable HTTP Server: The NCSA prototype", http://www.ncsa.uiuc.edu/InformationServers/Conferences/CERNwww94/www94.ncsa.html 3. R. Fielding, H. Frystyk, T. Berners-Lee, "Hypertext Transfer Protolcol - HTTP/1.1", Internet Draft, Jan. 1996 4. D. M. Kristol and L. Montulli, "Proposed HTTP State Management Mechanism", HTTP Working Group, Internet Draft, Feb. 1996 5. Netscape Corp, "Persistent Client State HTTPCookies" http://home.netscape.com/newsref/std/cookie_spec.html , Oct. 1995. 6. A. Freier, P. Karlton, P. Kocher, "The SSL Protocol, Ver 3.0", http://home.netscape.com/eng/ssl3/ssl-toc.html, Internet Draft, Mar. 1996 7. Zeus Technology Ltd, "ZeusBench/1.0", http://www.zeus.co.uk/products/server/intro/bench, 1996 8. P. J. Houston, "Transarc Corporation's World Wide Web Strategy Overview", http://www.transarc.com/afs/transarc.com/public/www/Public/ProdServ/WWW/l, 1996

Connection Caching for WWW Servers and Proxies
Session#: 620

Noritaka OSAWA
Graduate School of Information Systems
The University of Electro-Communications, JAPAN
osawa@is.uec.ac.jp

Kazuhisa YANAGAWA
Graduate School of Information Systems
The University of Electro-Communications, JAPAN
katze@yuba.is.uec.ac.jp

Toshitsugu YUBA
Graduate School of Information Systems
The University of Electro-Communications, JAPAN
yuba@is.uec.ac.jp

Abstract: This paper proposes connection caching to reduce the overhead of accessing WWW pages. Connection caching means that a WWW server or a WWW proxy server does not release its connection with a client or a peer but retains it and uses it again after the transmission is completed. The retained connection is cached and it is used for future access. Connection caching reduces network traffic and server load. This paper shows that the hit rate of connection caching for WWW accesses is high and connection caching is effective. This evaluation is based on actual logs from a server and a proxy server. The logs are records of 350 days. The log from the proxy server includes more than 14 million accesses.

1 Introduction

The World Wide Web (WWW) on the Internet is widely used. A large part of network traffic on the Internet is occupied by WWW accesses. User access latency and the server load have become problem areas. Sometimes users experience long access latency. The reasons are network congestion and server overload.

A caching technique is effective in solving these problems. By caching the frequently accessed data at sites near a client, network traffic and access latency decrease. Moreover, the WWW server load also decreases because the amount of accesses to the server is reduced by caching. Data caching has been used in servers, proxy servers and clients. A lot of research on data caching has been performed [Abrams et al. 1995] [Danzig et al. 1993] [Glassman 1994] [Ichii&Nakayama 1995] [Osawa et al. 1996] [Pitkow&Recker 1994]. We also proposed generational caching schemes for proxy server caches [Osawa et al. 1997].

Not only caching data but also caching of connections is possible. We propose connection caching for the WWW. It reduces network traffic and server load because the establishment of a connection creates some network traffic and needs server response. In this study, we will evaluate and discuss the hit rates of the connection cache, using access logs from the Information Processing Center (IPC) of the University of Electro-communications (UEC) in Japan. Our study of connection caching is based on logs of 350 days.

2 Evaluation Based on Logs

Servers where logs were gathered and our analysis method will be explained. The data that a Universal Resource Locator (URL) [Berners-Lee et al. 1994] refers to is called a page in this paper. In analysis of connection caching, the host part of URL is used to identify the peer.

2.1 Servers and Their Users

We will describe the proxy server and its users at IPC of UEC. After that, we will explain the WWW server.

Educational workstations (WS's) at IPC can not communicate directly with sites outside the university. Therefore users of educational WS's have to use a proxy server. IPC operates the CERN httpd [Luotonen&Altis 1994] as its proxy server. The proxy server is believed to be used by all users of educational WS's at IPC. The proxy server at IPC is also used as a cache by departments and laboratories that do not have their own proxy servers.

NCSA Mosaic [Andreessen 1993] is the WWW client on educational WS's at IPC. Mosaic used at IPC does not hold cached data beyond one session. On computers other than educational WS's at IPC, other WWW clients are also used in addition to Mosaic. Some clients, such as Netscape Navigator and Internet Explorer, hold cached data beyond one session.

The WWW server at UEC (http://www.uec.ac.jp/) is accessed from inside and outside UEC. It contains an introduction about UEC and has links to departments of UEC.

2.2 Logs

The relationship among clients, proxies and servers is shown in [Fig.1]. Client hosts access WWW pages on servers through proxy servers or directly. Logs from our proxy are used to identify both client host addresses and server host addresses. Logs from our WWW server are used to identify addresses of hosts which access pages on the server.

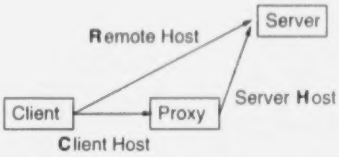


Figure 1: Relationship among clients, proxies and servers.

3 Characteristics of Log Data

We focus on successful accesses through Hypertext Transfer Protocol (HTTP) [Berners-Lee&Frystyk 1996]. Successful accesses will be referred to simply as accesses in this paper. Log data was gathered between 1 PM on October 24, 1995 and 1 PM on October 8, 1996. The length of log period is 350 days. The total number of accesses to the proxy was 14,270,689. That is an average of 40,773.4 access/day. The total number of accesses to the WWW server was 1,032,890. That is an average of 2,951.1 access/day.

[Fig.2] and [Fig.3] show access amounts by hour and by days of the week. A peak exists between 3PM and 4PM. Weekdays have more accesses than weekends. Distribution of accesses in [Fig.2] and [Fig.3] is as expected. There is nothing special about our logs.

[Glassman 1994] states that the access frequencies of pages in a WWW server follow Zipf's law [Knuth 1973]. If Zipf's law holds, there is locality of accesses. We investigated the access frequencies of hosts and

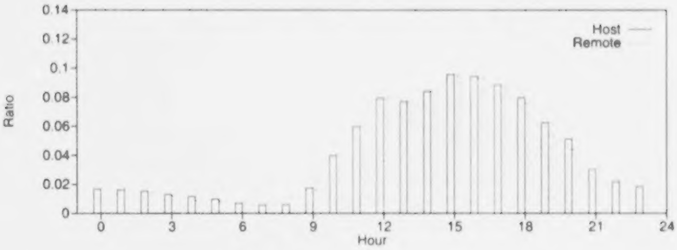


Figure 2: Normalized access frequencies by hour. Access frequencies are normalized by total number of accesses.

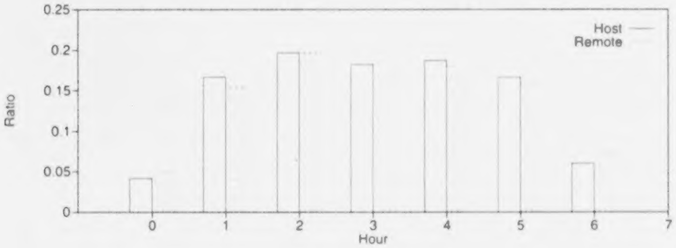


Figure 3: Normalized access frequencies by days of week. Access frequencies are normalized by total number of accesses. Zero and 6 in X-axis represent Sunday and Saturday respectively.

the frequencies of intervals between accesses to the same host. To our knowledge, this has never before been reported.

Let the number of pages whose frequency is f be $P(f)$. The number of page accesses of f is $A(f) = fP(f)$. If Zipf's law holds, $A(f) = M/f$ where M is a constant. By transforming equations, we get $P(f) = M/f^2$. $P(f)$ can be plotted as a line between top-left and bottom-right in log-log graphs like [Fig.4] and [Fig.5].

The number of hosts that have the same access frequency is shown in [Fig.4]. The X-axis is the access frequency of a host. The Y-axis is the number of hosts whose access frequencies are the same.

In [Fig.4], Zipf's law is applicable to the range of lower access frequencies of server hosts (Host) and remote hosts (Remote) but there are differences between the law and the log data. Access frequency of client host (Client) to the proxy does not match Zipf's law well. The number of clients of the proxy is limited because the clients must be hosts in the university.

The host that has highest access frequency is the WWW server host at IPC which have home pages related to the educational computers. Mosaic on an educational WS's accesses the home pages at startup time because it does not hold cached data beyond one session.

[Fig.5] shows the distribution of access intervals. The X-axis is the interval between accesses to the same host. An access interval is the interval between an access to a host and the next access to the same host. An access interval represents locality of accesses. High frequencies of small access intervals represent high locality of accesses. High locality improves the hit rate of a cache. The high hit rate reduces the overhead of accesses. [Fig.5] shows the high locality of accesses to hosts, thus connection caching would be effective. We will show quantitative evaluation of connection caching later.

In [Fig.5], Zipf's law is applicable to the range of smaller intervals between accesses to the same host.

BEST COPY AVAILABLE

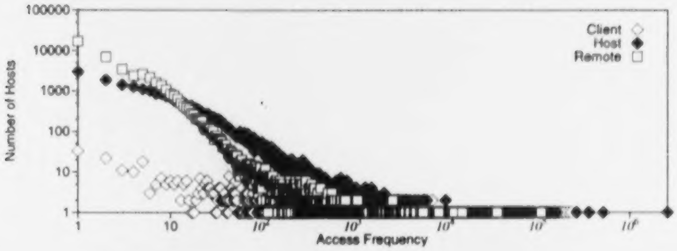


Figure 4: The number of hosts that have the same access frequency. Client, Host, Remote represent hit rates of client host at the proxy, server host at the proxy and remote host at the server, respectively.

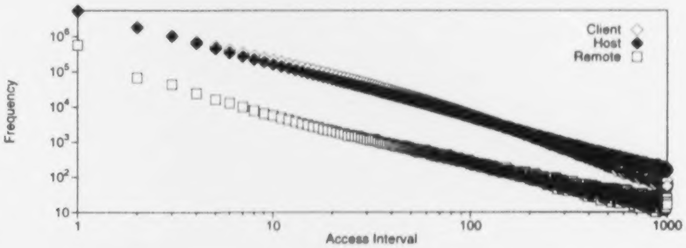


Figure 5: Distribution of access intervals. Client, Host, Remote represent hit rates of client host at the proxy, server host at the proxy and remote host at the server, respectively. This shows high locality of accesses.

4 Replacement Algorithms

We describe basic replacement algorithms for a cache. Conventional replacement algorithms [Maekawa et al. 1987] for a cache and their abbreviations will be explained.

LRU Least Recently Used algorithm. This algorithm replaces the least recently used entry with a new entry. LRU algorithm and simplified LRU algorithms are widely used in cache replacement.

FIFO First-In First-Out algorithm. This algorithm is used when a simple mechanism is preferred as hardware cache. This is easier to implement than LRU.

5 Caching of Connections

The evaluation of the caching of connections to clients (client hosts) on the proxy server, to servers (server hosts) on the proxy server, to clients or proxies (remote hosts) on the WWW server is shown in [Fig.6]. This evaluation is based on logs from the proxy server and the WWW server at IPC in UEC. [Fig.6] shows that connection caching is fairly effective.

The caching of 16 connections gives hit rates of more than 80% in all cases. The caching of 1024 connections gives hit rates of more than 90% in all cases. Hit rates using the LRU algorithm are superior to hit rates using FIFO as a replacement algorithm. However, the difference is not so large. Therefore the simple FIFO replacement algorithm is useful and effective in connection caching.

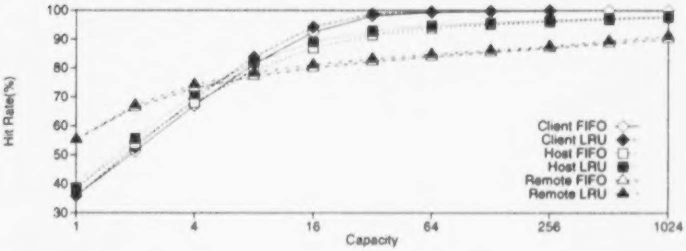


Figure 6: Hit rates of connection caching. Client, Host, Remote represent hit rates of client host at the proxy, server host at the proxy and remote host at the server, respectively. FIFO and LRU represent First-In First-Out replacement algorithm and Least Recently Used replacement algorithm respectively.

6 Discussion

TCP/IP connection is usually employed for WWW page access. Communications using TCP/IP are reliable and can pass network fire walls. Establishing a connection, however, is costly for servers and increases the access latency. Therefore it is important to investigate reduction of the overhead with connection-based protocols.

A protocol that transmits multiple pages in one connection was proposed in [Padmanabhan&Mogul 1994]. Pages that are referenced by the page that the user has accessed are transmitted together. That proposal aims at reducing access latency. It uses prefetching techniques. On the other hand, our proposal is based on caching techniques. Our proposal is different from [Padmanabhan&Mogul 1994]. Moreover we evaluated the effectiveness of connection caching quantitatively where the connection is not released after transmission of that page.

As shown above, connection caching is effective in all cases of proxy-from-client, proxy-to-server and server-from-hosts. However, this analysis assumes that only one connection is established for one host (server). When multiple accesses to a server are requested simultaneously, access latency may increase. An analysis of where multiple connections to a server are permitted is needed. Unfortunately, transfer time of pages is not recorded in the logs. Thus precise evaluation based on the logs is impossible. However, we can estimate the transfer time of pages on the basis of the sizes of pages and the average transmission rate. Hence we will study connection caching, where multiple connections are kept and access latency is minimized, based on estimations by varying the average transfer rate.

HTTP/1.1 [Fielding et al. 1997] introduced persistent connections, which lives until the connection is explicitly closed. With this facility, a connection caching scheme is more easily implemented than under the HTTP/1.0 environment, which has no standard for persistent connections. For example, a client side cache can be implemented simply by an array of persistent connections. The connection is closed when replacement occurs. The release of connections should be handled appropriately to operate the connection caching as an effective mechanism. Therefore a definite protocol should also be investigated.

7 Concluding Remarks

We proposed and analyzed connection caching for WWW severs and proxies on the basis of actual logs of 350 days. Our study showed that the caching of 1024 connections on a WWW server and a proxy server gives a hit rate of more than 90%. Connection caching utilizes the locality of accesses. Sufficiently large effectiveness is expected when using connection caching. LRU is superior to FIFO as a replacement algorithm for a cache, however, replacement algorithms do not affect the hit rates significantly. We have a plan to investigate and evaluate a definite protocol that uses connection caching.

Acknowledgments

We wish to acknowledge Prof. Shuichi Itoh, Dr. Katushi Kobayashi and other members of the Information Processing Center at the University of Electro-communications for helping us acquire the WWW access logs.

References

[Abrams et al. 1995] Abrams, M., Standridge, C.R., Abdulla, G., Williams, S., & Fox, E.A. (1995). Caching Proxies: Limitations and Potentials. *4th Int'l World Wide Web Conference.*

[Andreessen 1993] Andreessen, M. (1993). *NCSA Mosaic Technical Summary.* National Center for Supercomputing Applications.

[Berners-Lee&Frystyk 1996] Berners-Lee, T., Fielding R., & Frystyk, H. (1996). *Hypertext Transfer Protocol – HTTP/1.0.* HTTP Working Group Internet-Draft.

[Berners-Lee et al. 1994] Berners-Lee, T., Masinter L., & McCahill, M. (1994). *Uniform Resource Locators.* RFC 1738.

[Danzig et al. 1993] Danzig, P.B., Hall R.S., & Schwartz, M.F. (1993). A case for caching file objects inside internetworks. *ACM SIGCOMM, 1993*, 239–248.

[Fielding et al. 1997] Fielding, R., Gettys, J., Mogul, J., Frystyk, H., and Berners-Lee, T. (1997) *Hypertext Transfer Protocol – HTTP/1.1.* RFC 2068.

[Glassman 1994] Glassman, S. (1994). A Caching relay for the World Wide Web. *1st Int'l Conf. on WWW, 1994*, Geneva.; also appeared in (1994) *Computer Networks and ISDN Systems*, 27, 165–173.

[Ichii&Nakayama 1995] Ichii, S., & Nakayama, M. (1995). Effectiveness of WWW caching in a campus network. *IPSJ SIG Notes DSM-9505033.* (in Japanese)

[Knuth 1973] Knuth, D. (1973). *The Art of Computer Programming, Vol. 3. Sorting and Searching.* Addison-Wesley.

[Maekawa et al. 1987] Maekawa, M., Oldehoeft, A.E. & Oldehoeft R.R. (1987). *Operating Systems Advanced Concepts*, Benjamin/Cummings Pub., Menlo Park.

[Kwan et al. 1995] Kwan, T.T., MacGrath R.E., & Reed, D. (1995) . NCSA's World Wide Web Server: Design and Performance. *IEEE Computer*, 28(11), 68–74.

[Luotonen&Altis 1994] Luotonen, A., & Altis, K. (1994). World-Wide Web Proxies. *1st Int'l Conf. on WWW 1994*, Geneva.; also appeared in (1994) *Computer Networks and ISDN Systems*, 27, 147–154.

[Osawa et al. 1996] Osawa, N., Hayano, F., Yuba, T., & Hakozaki, K. (1996). Evaluation of Replacement Policies in WWW Cache Based on Logs from Proxy Serve r, *IPSJ SIG Notes DPS 74-33.* (in Japanese)

[Osawa et al. 1997] Osawa, N., Toshitsugu, Y., & Hakozaki, K. (1997). Generational Replacement Schemes for a WWW Caching Proxy Server. *High Performance Computing and Networking Europe (HPCN97)*, LNCS 1225, 940–949.

[Padmanabhan&Mogul 1994] Padmanabhan, V.N., & Mogul, J.C. (1994). Improving HTTP Latency. *2nd Int'l WWW Conf*, 995–1005.

[Pitkow&Recker 1994] Pitkow, J.E., & Recker, M.M. (1994). A Simple Yet Robust Caching Algorithm Based on Dynamic Access Patterns. *2nd Int'l WWW Conf.*, 1039–1046.

ASML: Automatic Site Markup Language

Charles B. Owen, Fillia Makedon, Glen Frank, Michael Kenyon
Dartmouth Experimental Visualization Laboratory (DEVLAB)
6211 Sudikoff Laboratory, Hanover, NH, 03755
http://devlab.dartmouth.edu/
devlab@cs.dartmouth.edu, (603) 646-3297

Abstract: Creation of large and complex World Wide Web sites is hampered by the "page at a time" approach of many tools and the programming knowledge and custom software development required for automated solutions. This paper describes the development of Automatic Site Markup Language (ASML), a new system designed to produce large and complicated web sites. ASML extends HTML with new, high-level features while still preserving complete compatibility with common browser and server technologies. It has powerful indexing and searching facilities, and enables the automatic translation of document formats. Most importantly, ASML provides HTML-like features at the site level rather than just the page level.

1. Introduction

Automatic Site Markup Language (ASML) is a new markup language designed to automate the construction of World Wide Web (WWW or web) sites. It centralizes functionality, decreases duplication of effort, and supplants most uses of scripting languages and custom programming in site development. It has already been utilized to construct a large historic site, The Prehistoric Archaeology of the Aegean [Rutter et al. 1996], as well as many smaller projects at the Dartmouth Experimental Visualization Laboratory (DEVLAB), including the ASML pages themselves (http://devlab.dartmouth.edu/asml/).

Web sites are created (authored) using several approaches. Content can be created by simply writing HTML manually using a text editor. Products also exist which allow for authoring pages with little or no knowledge of HTML in a form similar to working with a word processor. This "page-at-a-time" approach views a site as a collection of discrete hypertext documents. DEVLAB experience with web site development projects has revealed the shortcomings of the page-at-a-time approach. Creating the pages of a site one at a time is similar to creating a word processing document one page at a time. Were this report created this way, the authors would be forced to edit six page files wherever a change was made to overall document format.

One common authoring method is to "build" sites using custom scripts. *The DAGS'95 electronic conference proceedings* (http://www.cs.dartmouth.edu/dags/) were built using Apple Hypercard. *The Olympics in the Ancient Hellenic World* (http://devlab.dartmouth.edu/olympic/) was created using scripts based on *m4*, a macro language common in Unix systems, and Perl. These projects assumed a knowledgeable and experienced programmer and the need for *custom software development for every single project*. Most DEVLAB projects involve novice programmers as well as students not associated with Computer Science. ASML was developed to better support authoring in such an environment.

ASML incorporates many features which automate the development of sites. It can be thought of as a "markup language for markup languages." Figure 1 illustrates the use of ASML in a site. Site content originates in ASML, HTML, raw text, RTF and other formats. Static content served by the World Wide Web server consists of HTML pages, so ASML *incurs no performance penalty*. ASML also provides most automatic content generation and form data handling features which would otherwise require custom programming.

There have been many approaches to automating the development of web sites and to site-level authoring. An extensive list of related work is available in the full ASML technical report [Owen et al. 1997]. The scripting language approach is based on the creation of custom scripts which either edit an existing site (traversing all pages and changing a background image for example), generate a site from data or directly from a program, or generate content "on-the-fly" as CGI scripts. The most prevalent scripting language on the

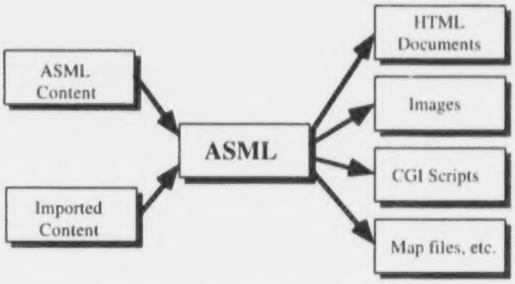


Figure 1 - Use of ASML at a site

WWW is the Perl language due to its wide availability. Other systems include Frontier from UserLand software, a popular Macintosh Environment (`http://www.scripting.com/frontier/`). Use of these systems is summed up by Christopher Hall and Carey Tews in a MacWeek review of Frontier: "Scripting in Frontier is by no means a job for beginners..." [Hall and Tews 1996].

Many graphical HTML editor tools are available. Carl Davis has compiled a list with reviews [Davis 1997]. Few editors support site-level features, the majority being designed for single page editing. Web Project Explorer from Haht Software incorporates a system of "clips", which are similar to ASML templates (http://www.haht.com/). However, clips are static macros which contain fixed content, as distinct from ASML templates which can be parameterized. Web Project Explorer has site level organization and visualization tools, but retains a page-at-a-time editing and composition approach.

There are many application areas where site-level authoring and the ability to evolve format are very important. On-line museums can involve huge amounts of material [Makedon et al. 1996]. Page-at-a-time authoring would be cumbersome and later revision (remodeling) time consuming. Many educators wish to use the WWW as a dissemination mechanism, leading to a large number of education specific approaches to site-level authoring and automatic content generation. ANDES Text Markup Language (ATML) is a markup language, much like ASML, but is specific to the ANDES distant education system [Johnson, Blake, and Shaw 1996]. CourseWeaver is an Apple Hypercard-based system designed for development of courseware for the web [Rebelsky 1997]. It uses a custom markup similar to that of ASML, provides for multi-targeted output, and handles content input translation.

2. ASML Capabilities

ASML has many powerful features which automate the generation of large WWW sites. This section describes some of these features. ASML is an integrated tool. All of the following categories of features are supported by a single environment and a single program, rather than a collection of disjoint tools.

2.1 Logical Modularization and Site-Level content generation

ASML abstracts a web site as a single hypertext document. The terminology *site* in this context refers to all content considered to be a single presentation, not necessarily all content on a specific server. As an example, the Prehistory Archaeology of the Aegean is considered a *site* even though it resides on the same server as several other major projects. Web sites have large amounts of intentional redundancy, usually enforcing consistent navigation methods and basic page appearance. This redundancy can be a major impediment to site reformatting. Ideally, site development would start by defining page formats, navigation, and site structure. However, content acquisition often proceeds in parallel with graphic development, parameters change as new content appears, and there's always someone who looks at the site and suggests a change after it is finished.

ASML allows common content to be defined as *templates*. Templates are referenced with a simple tag which *expands* to the common content. If a common element is changed, only the template definition need be changed. This is true for not only "tops and bottoms" of pages, the most common elements, but also for elements within pages such as graphical separators and local tables of contents.

ASML allows the generation of a complete site from a single ASML document. As in any large task, breaking the authoring task into smaller modules is very important. The modularization in traditional web-sites is *physical*, directly related to physical pages and groupings, typically in directories. ASML allows *logical modularization*, a division into whatever form best suites the user. An example might be grouping all common format elements in a single file, while a set of related pages group in another.

2.2 Multi-targeting, Location Independence, and Derivative Sites

ASML supports *multi-targeting*, providing content in more than one format. An example of multitargeting is providing sets of linked pages as well as a larger document suitable for printing or providing extended and condensed content. Another example is providing "frame" and "frame-free" content.

Location independence is the ability to produce a site at any required server address. This is a basic feature of ASML. Using the `<base>` tag, content can be relocated at will. In several DEVLAB projects this feature has been used to maintain simultaneous development and publication sites for the same project. The published site was updated occasionally when the content reached stable milestones. The development site contained content in incomplete stages.

The Prehistory Archaeology of the Aegean site is a *derivative site*, a site wherein the web publication is derived from content in an alien format. This is distinct from a *translated site*, wherein the web publication is a translation of content from an alien format. In this site content consists of a large set of notes by Prof. Jeremy Rutter of Dartmouth College. The source material is routinely expanded and revised and Prof. Rutter wishes to maintain it in the word processor format. ASML supports *imported content*, content in forms other than HTML or ASML (such as Rich Text Format, RTF, as used in this example), which is imported when the site is generated and used to construct the pages. The original document becomes a compilation source for the site. Whenever that document changes the site can be rebuild, incorporating the changes. This is important in any application where content routinely changes and must appear as both paper and web documents.

2.3 Content Searching, Indexing, and Simple CGI Scripting

Most larger web sites require indexing. Numerous tools are available for implementation of search engines at any level of a site. However, these tools are often complicated and not easily customized for a particular site. ASML has a built-in search engine which can be easily implemented using only ASML. The ASML search engine index file is constructed when the site is built and allows for very fast search operations. Search results are obtained quickly with minimum load on the web server. The search engine scores pages which match the query and ranks them on descending relevance. A page located using the search engine can be accessed with all search terms highlighted in red and the browser is automatically advanced to the first located search term.

One of the most complicated elements of site development is *Common Gateway Interface* (CGI) scripting. Scripting is required for processing data forms and dynamic content generation. Scripting usually involves programming, a specialized activity. ASML can serve as the scripting environment in many applications. Production of a script is little different from production of a page. The mechanism for acquisition of forms variables is automatic in ASML. All form variables are converted to templates. All issues of CGI protocol are managed by ASML, requiring no user knowledge of the underlying mechanisms.

3. Example ASML Sites

Several web sites have been created using ASML. This section describes two of these sites and discusses how ASML simplified site development. ASML is a general purpose system. It is not specific to any

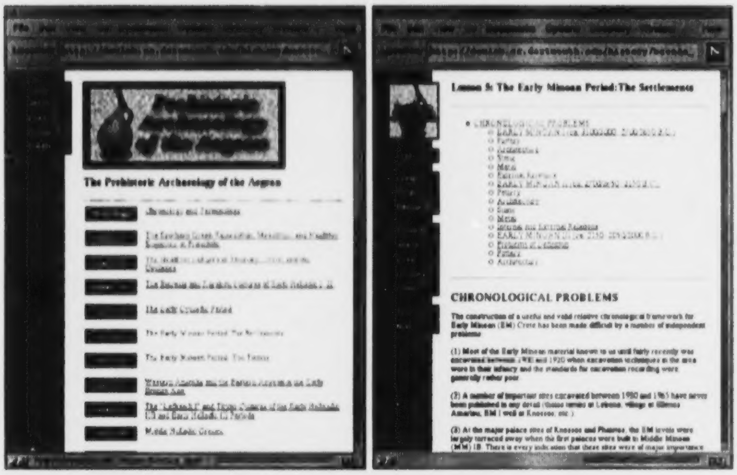

The Prehistoric Archaeology of the Aegean
Lesson 5: The Early Minoan Period:The Settlements
CHRONOLOGICAL PROBLEMS

Figure 2 - Prehistoric Archaeology of the Aegean site

application area. These two diverse applications illustrate this characteristic. In both cases ASML simplified design, but in no way forced any particular format.

3.1 The Prehistoric Archaeology of the Aegean

A major project at the DEVLAB has been the Prehistoric Archaeology of the Aegean [Rutter et al. 1996]. Figure 2 shows two pages from this site. This project and ASML development proceeded concurrently and the design of this site was a major influence on the design of ASML. The content developers for this site mostly consisted of students with no previous experience in WWW development. The site provides 29 lessens in archaeology and is designed for student use. Over 500 images of archaeological digs, historic locations, and significant artifacts are included in the site. All images are presented in a small browsing format, a larger display format, and the full high-resolution scan size.

This web site incorporates nearly all features of ASML. The site is a *derivative site* and can be reconstructed as the Microsoft Word content is updated (indeed, it has been updated several times). The ASML search engine is an integral feature of the site and four indices are generated. The pages which present the images are generated dynamically by an ASML CGI script. The page layout and navigation mechanisms were changed repeatedly in the process of site development. In spite of the fact that this site incorporates more than 100 pages, it is produced by 14 ASML files with a total of only 1759 lines.

3.2 ImageTcl Documentation

The ImageTcl multimedia development system has been developed at the DEVLAB for research in media data analysis [Owen and Makedon 1997]. The documentation for the system was migrated from HTML to ASML when it became available and is illustrated in Figure 3. The conversion to ASML was done in incremental stages. Initially ASML was used as a simple wrapper for HTML. Addition of templates to define common elements was done incrementally and has since become quite extensive. This is a common migration mechanism, illustrating that ASML can be added to an existing project with minimal effort. Each page includes a search engine query form. The table-based layout of the `imagefft` command parameters illustrated in Figure 3 is constructed using ASML templates, all user defined.

4. ASML Markup

BEST COPY AVAILABLE

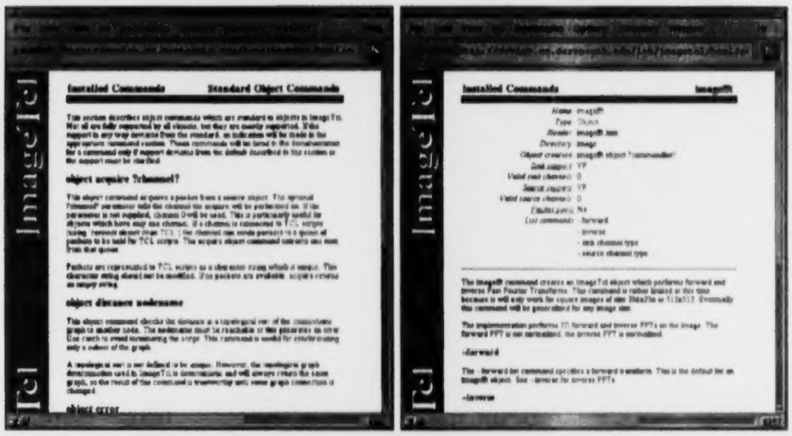


Figure 3 - ImageTcl documentation

Several important terms in ASML are *tag*, *end tag*, *attribute*, *value*, and *expand*. Markup in ASML (and HTML) is in the form of tags. A *tag* is the basic markup element. An *end tag* indicates the end of content considered to be contained by the tag. An *attribute* is an option on a tag which directs or enhances its functionality. A *value* is a string of information associated with an attribute. In both ASML and HTML, attributes are assigned values using the equal sign.

ASML *expands* tags when processing an ASML document. Expansion replaces the tag with different content. No ASML tags remain in the output of ASML execution (except when specifically added using the escape mechanism), so browsers need not be modified to support any ASML syntax. Indeed, ASML is invisible to the WWW user. The term expand indicates the processing of an ASML tag. Some ASML tags have conditional and repetitive processing functions and some expand to empty content.

A complete list of ASML tags is beyond the scope of this paper and is available as a technical report [Owen et al. 1997]. There are over two dozen basic tags and this set is easily expanded using the template mechanism. The tags can be grouped into several categories: template management (`{define}`, `{append}`, `{definelist}`), environment (`{base}`, `{include}`, `{page}`), CGI and form support (`{formget}`), searching and indexing (`{index}`, `{search}`, `{highlight}`), conditional and iterative execution (`{if}`, `{else}`, `{foreach}`), HTML enhancement (`{img}`), and derivative content (`{import}`, `{section}`).

4.1 Tag Format

The format for ASML tags and end tags is: `{tagname attribute=value attribute=value}` and `{/tagname}`. A fundamental design criterion for ASML is *familiarity for the HTML user*. Hence, the tag format is virtually identical to HTML. The only difference in notation is that curly braces { and } are used in place of the conventional < and > of HTML. The example below shows a typical HTML `<img>` tag and the corresponding ASML `{img}` tag:

```
<img src="/bronze/images/bar.gif" width=400 height=10 alt="---" align="center">
{img src="/images/bar.gif" alt="---" align="center"}
```

The use of an alternative markup delimiter allows intermingling of ASML and HTML tags while keeping the identity of each type of content intact. The HTML tag can still be used if desired. ASML is an extension of HTML rather than a replacement. ASML functionality can be easily added to an existing HTML document.

There are differences in the attribute values of these two tags. These differences are due to the *automation* incorporated in ASML. Image sizes are determined from the image, so there is no need to include height and width options in the ASML `{img}` tag. ASML is conscious of a home directory for a site (both on the server and on the local file system), so absolute addressing is relative to the site, not the server (in this case inside the `bronze` directory of the server). This allows a site to be location independent.

BEST COPY AVAILABLE

A tag can have 0 or more attributes. Values are usually contained in quotation marks, but these can be omitted if there are no spaces in the value or if the value consists only of another tag, a somewhat more liberal policy than in HTML. ASML can contain tags within tag names or attribute values. The following is a valid ASML tag: `{lesson-{lesson}-text}`. The inner `{lesson}` tag will expand first to create the name of the outer tag. In this example, `{lesson}` might expand to the lesson number and `{lesson-1-text}` expand to the text for lesson one. The ability to produce multiple pages from imported content is valuable in courseware production.

Comment sequences begin with `{--` and end with `--}`. All content between the start and end of a comment sequence is ignored by ASML and are not translated to HTML comments or reproduced in any way in the output page. HTML comments are often underutilized due to the fact that any user can view the source of a page and see the comments and they increase the transmitted file size. ASML comments are fully available to the site author, but not the casual user.

4.2 Templates

A fundamental design element of ASML is the *template*. A template is text, possibility with optional "fill-in" content, which is accessed using an ASML tag. Templates can define components of pages which are common throughout the site. Templates also have the power to build tables of contents and other cumulative components. Each tag in an ASML document is tested against the standard *system tags*. If the tag is not a system tag, it is tested to see if it is a template and expanded to the template contents. Template expansion in ASML continues until there are no ASML tags remaining. Templates can contain tags which will, themselves, be expanded. ASML defers expansion of templates, so templates with tags are defined as including the tags, *not the expansion of the tags*. Expansion takes place where the template is used. Templates in ASML can be used in pages, other templates, and even tag names and attribute values. Deferred expansion can be overridden.

4.3 Consequences of the ASML Markup Format

The markup format used in ASML was a difficult decision. Many alternatives were examined such as the `<@tag>` structure used in ATML [Johnson, Blake, and Shaw 1996]. However, it was felt that that format would be difficult to distinguish from HTML and would be confusing for novice page authors. The use of braces is not ideal in that it conflicts with the format proposed for *cascading style sheets*, a proposed format specification mechanism which has been adopted by some browsers [Lie and Bos 1996]. To set the text color of the "H1" elements of an HTML document to blue, the following style sheet entry would be used: `H1 { color: blue}`. The braces can be included using the escape mechanism in ASML: `H1 \{ color: blue\}`. When ASML was under development the cascading style sheet proposal had been considered dropped and has only recently been revived. Several syntactic modifications for simplifying this problem are being examined including structures similar to the extended quote in the Perl language or inclusion of raw content from additional files. The ASML markup format also conflicts with JavaScript. Most of the same issues apply in that case.

5. Conclusion and Future Work

ASML,a new approach to site-level World Wide Web development, views a site as a complete document, not a collection of disjoint pages. It retains the simple structure of HTML and total compatibility with existing browsers and servers while providing new capabilities such as centralization of common page elements, search and indexing features, and data import. The syntax is similar to HTML and, therefore, easy to learn and use. The environment is not fundamentally a programming environment and no programming skills are required to use ASML. In many cases ASML can replace programming solutions in site development.

ASML is currently in version 1.03 and many new features are planned. Most HTML tags, especially the `<a>` tag, will be duplicated in ASML, giving the system more power to direct the page generation process. The current structure of ASML builds new sites whenever `asml` is invoked. While ASML is very fast, it may be inefficient to generate a large site in this way. A system of automatic dependency checking is planned which will allow only pages with changed content to be rebuilt.

Bibliography

[Davis 1997] C. Davis (1997), *Carl Davis's HTML Editor Reviews*, `<http://homepage.interaccess.com/~cdavis/editrev/>`.

[Hall and Tews 1996] C. Hall and C. Tews (1996), UserLand's scripting system fulfills its manifest destiny, *MacWeek*, 10(28), July 22, 1996.

[Johnson, Blake, and Shaw 1996] W. L. Johnson, T. Blake, and E. Shaw (1996). Automated management and delivery of distance courseware. In *Proceedings of WebNet'96 World Conference of the Web Society*, San Francisco, CA. 225-230.

[Lie and Bos 1996] H. W. Lie and B. Bos (1996), *Cascading Style Sheets, level 1*, W3C Recommendation REC-CSS1-961217, December 17, 1996, `<http://www.w3.org/pub/WWW/TR/REC-CSS1>`.

[Makedon et al. 1996] F. Makedon, J. Ford, M. Kenyon, and C. Owen (1996). Ancient museum collections and the web. In *Proceedings of WebNet'96 World Conference of the Web Society*, San Francisco, CA, 1996. 315-329.

[Owen and Makedon 1997] C. B. Owen and F. Makedon (1997), Multimedia data analysis using ImageTcl. *Gesellschaft für Klassifikation e.V.*, University of Potsdam, Potsdam, Germany, March 12-14, 1997, to appear.

[Owen et al. 1997] C. B. Owen, F. Makedon, G. Frank, and M. Kenyon (1997), *ASML: Automatic Site Markup Language 1.03*, Dartmouth College Technical Report PCS-TR97-309.

[Rebelsky 1997] S. A. Rebelsky (1997), CourseWeaver: A tool for building course-based webs, *Proceedings of ED-MEDIA'97*, Calgary, Canada, June 14-19, 1997.

[Rutter et al. 1996] J. B. Rutter, et. al. (1996), *The Prehistoric Archaeology of the Aegean*, `<http://devlab.dartmouth.edu/history/bronze_age/>`.

Virtual Partnerships in Research and Education

Deborah A. Payne, Kelly A. Keating, James D. Myers
Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory,
K8-91, P.O. Box 999, Richland, WA 99352, USA
da_payne@pnl.gov, kelly.keating@pnl.gov, jd_myers@pnl.gov

Abstract:
The William R. Wiley Environmental Molecular Sciences Laboratory (EMSL) at the Pacific Northwest National Laboratory is a collaborative user facility with many unique scientific capabilities. The EMSL expects to support many of its remote users and collaborators by electronic means and is creating a collaborative environment for this purpose with capabilities ranging from chat and video-conferencing, to shared applications, electronic notebooks, and remote-controlled instruments. This paper describes some of the particular capabilities required to support scientific collaborations, the status and direction of the EMSL tools, and several early uses of the EMSL software in both research and education collaborations. Together, these topics define a vision for natural, in-depth, virtual partnerships in research and education.

Introduction

The move toward virtual enterprises, seen in today's business world, is also occurring in the field of scientific research and education. National laboratories, such as the Pacific Northwest National Laboratory (PNNL), are making their data, instruments, and expertise, available to academic, industrial, and government collaborators, and conversely, are planning to make use of physical and intellectual resources at other institutions to supplement their own capabilities[WWW 97a][Kouzes et. al. 96]. Educators are looking to provide students with training in the latest techniques using state-of-the-art instrumentation as well as to motivate students' learning with cross-disciplinary examples of the use of science knowledge to solve real-world problems. PNNL's Environmental Molecular Sciences Laboratory (EMSL)[WWW 97b] is a new $230M facility for basic research in environmental and molecular sciences in support of the Department of Energy's mission to develop new technologies to clean up the nation's hazardous waste sites. The EMSL will house many unique facilities for basic scientific research, including the world's first commercial near-gigahertz Nuclear Magnetic Resonance (NMR) spectrometer, a scanning near field optical microscope, and the most powerful IBM parallel supercomputer yet built. Overall, the EMSL will house nearly 300 researchers with unique expertise, equipment, and software, seeking to understand the fundamental physical, chemical, and biological processes that underlie:

- the use of natural and engineered techniques to remediate and restore contaminated soils and groundwater,
- the processes and techniques used to extract and destroy chemical wastes, and to separate and safely store radioactive wastes,
- the development of a new generation of industrial processes that minimize or eliminate the use of toxic materials and the production of hazardous waste products, and
- the impact of toxic contaminants on the health of humans and the ecosystem.

While there are many research and education collaborations today involving EMSL researchers and their remote colleagues, the use of electronic collaboration tools promises to greatly enhance both the quantity and quality of such collaborations. However, simple videoconferencing tools are not sufficient to allow natural, in-depth collaboration, especially when the topic involves a complex scientific instrument, exploring multi-dimensional data, or synthesizing results from theory and experiment. Through interviews with EMSL researchers and their colleagues, and iterative feedback during software development, we've developed a simple taxonomy to describe the types of research collaborations that currently exist and have evaluated the communications needs in each type.

Some collaborations involve researchers in the same field sharing an instrument. The remote researcher might contribute to the design of a new detector and then use the instrument to study molecular systems of interest. In this **peer-to-peer** type of collaboration, the researchers share a common scientific vocabulary. The most important aspects of their collaborations are shared instruments and unanalyzed data, making remote instrument control and direct data file access important.

Other collaborations involve senior scientists and their more junior partners, such as students and postdoctoral fellows. In these collaborations, the mentor may use prepared materials and live demonstrations to teach data acquisition, analysis techniques, and scientific principles. The mentor must then observe as the student demonstrates mastery of the new concepts by using them appropriately. The necessary real-time interactions between mentor and student go far beyond standard conferencing: a mentor and student must be able to work collaboratively and interactively, sharing a view of an experiment in progress or the live output of a modeling/visualization package. In this **mentor-student** type of collaboration, real-time interactions are supplemented by asynchronous access to many types of archival information - data, notes, results, etc. This also allows the student to revisit the material as needed.

A third type of collaboration anticipated is between scientists doing complementary studies of the same molecular systems. For instance, a theorist may calculate structures of molecular clusters while an experimentalist uses laser spectroscopy to make an experimental measurement of the structure. Researchers in such **inter-disciplinary** collaborations share less of a common vocabulary and must often translate their results into each other's terms, alternating between the roles of mentor and student. Direct access to instruments or to raw data becomes less useful to the researchers, while access to summaries and analyses, perhaps recorded in an electronic notebook, and the ability to discuss unfamiliar concepts and to correct misunderstandings become more important.

A fourth type of collaboration, again involving researchers in different disciplines, involves one researcher, or research team, providing input for another. Examples of this type of collaboration include a mass spectroscopist determining the sequence of a protein or other biopolymer for a biologist, or a surface scientist providing reaction rate data to a geologist modeling the subsurface transport of hazardous wastes. Working with an analytical laboratory on a fee-per-service basis represents an extreme form of this **producer-consumer** type of collaboration. There is often a wider gap between the disciplines and motivations of researchers in such collaborations; a scientist may be interested in a new physical phenomenon while their collaborator, an engineer, is trying to reduce the cost of a clean up effort. They may have little chance for professional contact in their daily work or at conferences. Researchers in these types of relationships place the strongest emphasis on being able to receive a sample and information about it, and being able to transmit results back to the other party. However, new ideas and approaches can appear if these researchers communicate more closely. The EMSL and PNNL hold seminar series, workshops, and pizza dinner discussions, to foster this type of communication between basic and applied scientists. This suggests that if these researchers are provided with readily available tools for informal electronic discussions, their collaboration may become more complementary as they adjust their studies to incorporate new ideas from each other.

These collaboration types suggest a range of useful tools, from email, voice and video, to shared computer displays, remote instruments, and electronic notebooks. Different collaboration types, and different tasks within them, will stress different communications channels. Other aspects of collaboration, and scientific collaboration in particular, affect the design of a shared electronic environment. During any collaboration, communication naturally switches between media as appropriate. An electronic collaboration environment should allow someone to talk, shrug, draw a graph, and point at new data from an instrument, all with minimal awareness of having switched to a new tool. Similarly, collaborations may move through different phases - acquiring data, analyzing results, writing papers - that require different communication tools. A collaboration environment must support easy transition between tools as required. Lastly, most scientific collaborators have intermittent contact. A relationship may lie dormant for weeks or months and then enter a period of high activity after a new capability is developed or new data is obtained. Any collaboration environment must support this use pattern.

It is important to note that while these classifications and examples all relate to scientific research, similar collaborations arise in education and business. Students may ask professors for help while working in teams of peers on projects. Workers might have peer-to-peer collaborations within their organization, and mentor-student or producer-consumer collaborations with suppliers and customers. Thus, software that is designed to support scientific collaborations will be applicable in other domains as well.

The EMSL Collaboratory Tools

EMSL's real-time Collaborative Research Environment (CORE) provides users with a single, simple way to access multiple electronic collaboration capabilities independent of their computer platform. CORE has a World Wide Web (WWW) main interface and provides cross platform capabilities to the user via both new software developed for CORE and via existing stand-alone tools, or combinations of compatible tools, that have been integrated. CORE hides the different syntax each tool has for launching and connecting to collaborators, helping to make collaboration more natural. Users start and join sessions using their names and a short topic description. Computer addresses, port numbers, and firewalls: all disappear from the user's view.

CORE relies on a central session manager and desktop executives that coordinate communications between participants and configure the various collaborative components. Use of the WWW paradigm makes the system easy for users to understand. The main interface of CORE is a WWW page that allows users to start or join collaborative sessions via a WWW form. This page, shown in [Fig. 1a], uses a common gateway interface (CGI) script to process user input. To start a new session, the user enters their name in the "User Name" text box and a brief topic description in the "Session Name" text box, and clicks on the "Start a New Session" push button. To join an existing session, the user enters their name and clicks the button showing the desired topic in the "Active Sessions" list.

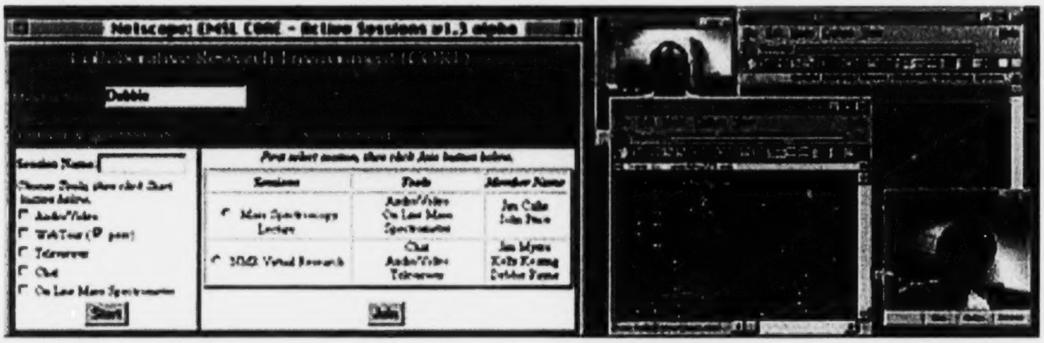


Figure 1. a) CORE's simple WWW user interface: select tools and start a session or click to join an existing one. b) Researchers discuss NMR data via CORE

When a new session is started, the user may select the tools desired for the given session. The session manager may start server processes for some of the tools, such as the EMSL TeleViewer described below. For other tools, such as videoconferencing, the user's IP address and platform type are used to determine the appropriate parameters for launching the client videoconferencing software. In our environment, we have implemented two third party options for audio/video conferencing. One is Cu-SeeMe, which we have implementing using a CU-SeeMe[WWW 97d] reflector bridge across PNNL's firewall. Macintosh and PC users use CU-SeeMe, connected to the appropriate end of the bridge to conference. The second option is use of multi-cast MBone[Eriksson 94] tools. Unix and PC users can launch MBone audio/video to either run independently or to connect to the reflector bridge. The session manager determines the appropriate parameters to launch software on each user's machine.

Once all the connection information is determined, and appropriate servers are started, the CGI script sends a custom multipurpose internet mail extension (MIME) typed file to the user's browser. The CORE desktop executive is started as a viewer (helper application) for this custom MIME file, just as a video player is started

BEST COPY AVAILABLE

to "view" a video/mpeg MIME type movie file. The helper application was developed in Java as is the session manager. The executive prepares its own communications, either opening a listening socket, or connecting to a listening executive, and then launches the requested collaborative tools. CORE provides a basic set of tools, some of which have been developed as part of the Collaboratory project and are highly integrated with the CORE executive, and others that are the product of other EMSL projects and third party efforts and use their own communications once launched. A brief description of each of the capabilities follows:

A. WebTour: WebTour provides the ability to synchronize WWW browsers, allowing users to hold lectures or discussions, using material on the WWW. WebTour can be run in either lecture mode (only the leader's browser is echoed) or peer-to-peer modes. The WebTour functionality is embedded in the CORE executive and uses its communications to the browser and to other executives.

B. File Sharing: The CORE provides file sharing as an extension to the WebTour. Any local files opened in the user's WWW browser are transmitted to collaborators and opened with their browsers. Because it uses the WWW's browser/viewer mechanism, it allows remote users to choose different applications to view transferred files, i.e. users may choose different word processors to view a rich text format (RTF) file.

C. Chat Box: A simple chat box is included in the executive as well. Messages are tagged with the user names given when starting the session. Proper serialization is guaranteed by sending all messages to the central executive (the one that started the session) which then redistributes them to all executives in the session.

D. TeleViewer: The EMSL TeleViewer[Keller 96] provides a cross platform shared computer display. Users may select a rectangle or window from their computer, or their entire display to share with collaborators. Using this tool, users can view any program running on the shared display, such as word processors, spreadsheets, instrument control software, and mathematical computations. The TeleViewer will soon provide annotation on top of the live image and eventually the ability to remotely control the shared application.

E. Electronic Notebook: The EMSL Electronic Laboratory Notebook (ELN)[Myers 96] provides users with a shared, interactive version of the traditional paper laboratory notebook. The current system allows users to create secure, dynamic, searchable, WWW pages, organized in notebooks, with text, links, images (files or screen capture), live views of the data with information about each file (instrument parameters that were used, the operator's name, the date, etc.), etc. The notebook is easily extended to handle additional data types. For instance, we recently added the capability to view protein structures stored in the protein data bank (pdb) format by incorporating a third party Java applet. Data from EMSL instruments can be sent directly to the ELN, where it is immediately available for viewing, download, comment, and analysis by all collaborators.

F. On-line Instruments: Other projects within the EMSL are developing on-line instruments that can be run remotely via the internet. CORE provides a mechanism to select and launch this software as part of a real-time session, while the notebook provides remote access to the acquired data and other information. One of the first of these instruments is a remote enabled radio frequency ion trap mass spectrometer. Commercial instruments, such as the EMSL's Varian Nuclear Magnetic Resonance (NMR) spectrometers, which already have remote capabilities, are being integrated with CORE and the electronic notebook.

G. Whiteboard: Whiteboards provide a shared space where users can write and draw, on a blank canvas or over a preexisting image.

H. Audio/video conferencing: Audio/video conferencing allows collaborators to see and hear each other, as well as to monitor instruments and laboratories. CORE currently launches CU-SeeMe or MBone's 'vic' and 'vat', depending on the user's preference and platform. As part of the Collaboratory project, PNNL set up a CU-SeeMe reflector bridge across our firewall that allows conferencing between EMSL researchers and external colleagues, while managing security.

Collaboratory Use

CORE and/or the electronic notebook are being used by several groups, most of which consist of an EMSL researcher or group working with their remote colleagues, though some groups have no EMSL research connection. The interests of the groups range from software development to quantum chemistry, mass spectroscopy, NMR spectroscopy, and reactive transport modeling. Some groups are strictly research oriented while others are using the collaborative tools to provide student research opportunities or to bring instruments and remote experts into classrooms. There have also been many non-science demonstration and trial uses of CORE, ranging from business meetings/presentations, remote training, and rapid response intelligence analysis.

The two groups described below demonstrate the use of CORE and the notebook in research and education settings.

An NMR Virtual Research Facility project is using the Collaboratory tools to let PNNL and Lawrence Berkeley National Laboratory (LBNL) structural biology researchers work closely together to determine the solution structure of proteins and DNA molecules. The NMR data will be collected at PNNL on a Varian 750 MHz NMR spectrometer, a resource not available at LBNL. Once the sample is inserted into the probe by a PNNL researcher, experiments can be run locally, or remotely and securely via the internet. In the initial joint experiment, between Dr. Kelly Keating of PNNL and Dr. Jeff Pelton of LBNL, preliminary work included sharing background references, known structures of similar molecules, and project plans through their electronic notebook. Dr. Pelton learned the specifics of PNNL's 750 MHz spectrometer control software by virtually sitting in on one of Dr. Keating's experiments, viewing the spectrometer console in real time via the Televiewer, while discussing the experiment parameters via videoconferencing, and recording notes in the electronic notebook. For the first, and several subsequent data acquisition sessions, Dr. Pelton controlled the experiment remotely with Dr. Keating observing. During experiment runs, which could last for two days for two dimensional (2D) NMR spectra, either or both collaborators would log directly into the spectrometer to check the progress, and/or use CORE to share the progress report and discuss the experiment. The notebook allowed similar, asynchronous, discussions, with Drs. Keating and Pelton viewing and commenting on current 2D data slices posted to their shared notebook. Once data were acquired, the collaborators continued to use CORE and the notebook as they began processing the data and assigning signals to specific atoms in the molecule. During analysis, the notebook again allows each researcher to link a copy of their results to the relevant ELN page with screen snapshots and comments to guide the other's work. Collaborative sessions, using videoconferencing and the TeleViewer allow joint analysis to complete difficult assignments. After perhaps several months of this cycle of NMR data collection, accessing the data, and data analysis, the collaborators will begin to jointly write their results into a paper for publication, exchanging documents and figures via the notebook and discussing changes on the fly using CORE.

The Collaboratory tools have also been used to provide a remote lecture to Professor Jim Callis' Chemistry 155 class at the University of Washington. The students were given a quick mass spectroscopy tutorial via videoconference and the WebTour by Dr. John Price at the EMSL, and then used his ion trap mass spectrometer remotely to complete a laboratory assignment, comparing the calculated and experimental spectra of a molecule containing three chlorine atoms. Their data was instantly available to all participants via the WWW based notebook. The ion trap mass spectrometer and CORE tools have also been used in student research collaborations with the University of Washington and Heritage College. In these cases, students were able to participate remotely, over a long term, in publishable research projects involving their local advisors and EMSL researchers.

Conclusions

The Collaboratory has been developed over the past two years, a time in which the internet and the WWW have changed greatly. In particular, the emergence of Java and distributed object frameworks, such as the Common Object Request Broker Architecture (CORBA), promise to revolutionize the development of dynamic WWW interfaces. As the current generation of CORE and the ELN move into productive use by EMSL researchers, we are also moving into a new round of development, in concert with other national laboratories, as part of the DOE's DOE2000 Collaboratory project[WWW 97d]. DOE2000 will result in a very usable set of tools as well as an extensible architecture for the development of more advanced and more domain specialized collaborative tools. (DOE2000 also includes two pilot collaborations with distributed academic, government, and industrial participants that will use the tools to enhance their research.) An iterative development approach will help ensure that the Collaboratory tools will meet the needs of collaborating researchers. To provide a similar experience in educational use of collaboration technologies to link national laboratories and academic sites, the EMSL and eight northwest academic institutions have formed the Collaboratory for Undergraduate Research and Education (CURE) group [Myers et. al. 97]. This NSF and DOE funded group is developing ways to maximize the benefit to students of exposure to the data, instruments, and expertise of the laboratories through combinations of remote lectures and laboratory experiments, student research projects, faculty development, etc. A major goal of the project is to encourage a web of collaboration between academic sites and the laboratory that will scale much better than a multitude of one-to-one collaborations.

Collaborative environments, such as the EMSL's Collaboratory suite, can provide users with a powerful array of collaborative capabilities to support distributed scientific research and education collaborations. By hiding the complexities of configuring individual tools, and providing cross-platform capabilities, collaborative environments reduce the barriers to communicating with remote colleagues. Extensions to the standard videoconferencing tools such as the Televiewer shared computer display, remote instruments, and electronic notebooks, allow collaborators to bring scientific resources directly into their discussions. Such environments hold the promise of making work with remote colleagues as simple, natural, and effective as working with people down the hall.

Acknowledgements

This work was supported by the U. S. Department of Energy through the DOE2000 program and the Distributed Collaboratory Experiment Environments (DCEE) program, both sponsored by the Mathematical, Information and Computational Sciences Division of the Office of Energy Research, and through the Laboratory Directed Research and Development program at Pacific Northwest National Laboratory. Pacific Northwest National Laboratory is a multiprogram national laboratory operated by Battelle Memorial Institute for the U.S. Department of Energy under Contract DE-AC06-76RLO 1830. We gratefully acknowledge the contribution of many individuals to this work (see [WWW 97a]).

References

[Eriksson 94] *MBONE, The Multicast Backbone*, H. Eriksson (1994). Comm. ACM, Aug. 1994, pp. 54-60, also see http://www.mbone.com/mbone/mbone.faq.html

[Keller 96] *The EMSL TeleViewer: A Collaborative Shared Computer Display*, P.E. Keller, J.D. Myers (1996). IEEE Fifth Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, June 19-21, 1996, Stanford University, Palo Alto, CA

http://www.emsl.pnl.gov:2080/docs/collab/papers/televiewer.wetice.html
[Kouzes et. al. 96] R.T. Kouzes, J.D. Myers, W.A. Wulf (1996*). Collaboratories: Doing Science on the Internet*, IEEE Computer, 29:8.

[Myers et. al. 96] *Electronic Laboratory Notebooks for Collaborative Research*, J.D. Myers, D. Le, J. Laird, C. Fox-Dobbs, D. Reich, T. Curtz (1996). IEEE Fifth Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises, June 19-21, 1996, Stanford University, Palo Alto, CA http://www.emsl.pnl.gov:2080/docs/collab/papers/notebook.wetice96.html

[Myers et. al. 97] *Collaboratories: Bringing National Laboratories into the Undergraduate Classroom and Laboratory via the Internet*, Jim Myers, Norman Chonacky, Thom Dunning, Eric Leber (1997), Council on Undergraduate Research Quarterly, March 1997

[WWW 97a] *Collaboratory for Environmental Molecular Sciences*, Pacific Northwest National Laboratory http://www.emsl.pnl.gov:2080/docs/collab/

[WWW 97b] *W. R. Wiley Environmental Molecular Sciences Laboratory*, Pacific Northwest National Laboratory http://www.emsl.pnl.gov/

[WWW 97c] *CU-SeeMe Welcome Page*, Cornell University, http://cu-seeme.cornell.edu/

[WWW 97d] *DOE2000 Project Homepage*, http://www.mcs.anl.gov/DOE2000

Telemedicine: An Inquiry in the Economicand Social Dynamics of Communications Technologies in the Medical Field

Francis Pereira, Ph.D., Center for Telecommunications Management, University of Southern California
Elizabeth Fife, Ph.D. Candidate, Center for Telecommunications Management, University of Southern California
Antonio A. Schuh, Center for Telecommunications Management, University of Southern California

ABSTRACT

The heath care industry in the United States has been experiencing substantial and ever-increasing cost pressures. Telemedicine, in this respect, offers significant potential for addressing some of the challenges faced by the health care industry. However, despite the fact that Telemedicine technology has existed since the 1920s, its use has not been widespread. The use of the diffusion of innovation theory as an organizing framework, coupled with results of a survey of Telemedicine professionals at the Global Telemedicine 2000 Conference in Chicago in 1996, identifies Telemedicine's potential as well as the barriers that are impeding its wide-spread application. These barriers include several social constraints, particularly: i) low compatibility with existing medical practices; ii), complexity of Telemedicine equipment and interfaces; iii) absence of reimbursement by third party agencies; and incompatibility of state laws regarding Telemedicine and licensure issues.

Introduction

Although aggregate and per capita costs of health care in the United States are the highest in the world, many Americans still remain uninsured, under-insured or live in communities that are medically under-served. A recent report from the Health Care Financing Agency (HCFA) estimates that annual health care expenditures exceed $900 billion, which amounts to more than $2 billion dollars a day or the equivalent of almost 15% of the Gross Domestic Product (Clyburn, 1996). In sharp contrast, it is estimated that some 15% to 25% of Americans live in counties that are defined as medically under-served (Office of Technology Assessment, 1990). Equally important between 1980 and 1989, the costs of medical services increased by 99%, or at twice the rate of inflation during the same period (National Telecommunications and Information Admistration, 1991). Furthermore, in 1994, the 4.8% increase in medical costs still represented more than twice the overall rate of inflation of 2.3% and exceeded the increase in workers' earnings of 2.5% (Swartz, 1994).

In this regard, Telemedicine, generally defined as "the use of telecommunications and computer technologies with medical expertise to facilitate health care delivery" (Kim, Cabral & Kim, 1995), has significant potential for developing into an integral component of the global health care system. Through remote sensing, collaborative patient care and access to electronic libraries and medical databases (Lindberg, 1994), Telemedicine can engender better and more extensive access to health care, lower medical costs, reduce the isolation of medical care professionals and increase medical productivity.

Although Telemedicine has existed since the 1920s [1] (Williams and Moore, 1995), it thus far, has been used only sparingly for real-world patient-physician consultations. A study conducted by Abt Associates found that, even when a broad definition was used, only 18% of all rural hospitals in the US were using Telemedicine. Furthermore, there has been a very limited number of clinical studies documenting Telemedicine's efficacy as a primary diagnostic and treatment tool (Perednia and Allen, 1995). Rigorous technology assessments that could form the basis for a coherent guide to the cost effective use of integrated systems are also lacking.

Although telemedicine offers significant advantages, its limited use suggests a lack of compatability with existing experiences and values. The use of the Diffusion of Innovation Theory, as an organizing framework, helps elucidate the benefits that Telemedicine offers to potential adopters and identifies the barriers to the increased and widespread use of Telemedicine.

---

[1] A form of telemedicine was used in the 1920s, when radio was used to link public health physicians standing at watch at shore stations in order to assist ships at sea that had medical emergencies. In the late 1950s, attention was drawn to closed circuit [*television*] systems using microwaves (Kim, Cabral, Parsons et alli, 1995), and in the 1970s satellites were used in large demonstration projects linking Alaskan and Canadian villages under the auspices of the NASA.

Telemedicine under the Diffusion of Innovations Framework

Some innovations such as pocket calculators or camcorders diffuse from first introduction to widespread use, or critical mass, within a few years. Others, like Telemedicine, require a longer time. Several models can be used to explain the differences in the rate of adoption. Generally, these models dichotomize members of the social system into early adopters and late adopters.

Late adopters either observe and imitate early adopters, or they communicate with them and are persuaded or induced to adopt these services, products or technologies, and critical mass is eventually achieved. One such model for these processes, the "diffusion of innovation theory," (Rogers, 1995) suggests five characteristics which can be used to describe innovations and analyzes how individuals' perceptions of these characteristics affect the adoption rate. These are summarized below:

Relative Advantage	The social and economic advantages that can be derived from adopting the new product
Reduction of Uncertainty	*Compatibility*: the degree to which an innovation is perceived as consistent with existing values and past experiences of the adopter. *Complexity*: the extent to which the innovation is perceived as difficult to understand and use. *Trialability*: the degree to which the innovation can experimented with on a limited basis. *observability*: the degree to which the results of an innovation are visible to others.
Social System	nature of the social system, which is the set of interrelated units engaged in joint problem solving, its structure (formal and informal) and its norms.
Type of Innovation-Decision	Optional-based or authority/consensus-based decision making.
Communication Channels	extent of change agents' promotion efforts where change agents are opinion leaders who could influence other members of the social system to adopt (or conversely not adopt) an innovation

Relative Advantage

Economic Advantages:

Although there have been no definitive cost-benefit analyses to determine the economic viability of Telemedicine projects, several studies have demonstrated the cost saving potential of Telemedicine. For example, a study prepared by the Arthur D. Little consulting company estimated the benefits at $36 billion annually (Moore, 1995). These savings could be generated from: (i) *reduced costs for serving patients*, through savings in time and travel for doctors and patients, fewer unnecessary referrals, and the replacement of doctors with less medically trained personnel supported by Telemedicine (Moore, 1995); ii) *cost savings from the provision of better health care*, generating cost reductions from early diagnosis and treatment. The cost saving benefits of Telemedicine have been substantiated in several studies, including the case of Texas Tech MEDNET which demonstrated savings of $1000 per patient when the patient was locally treated (Williams and Moore).

Currently, however, for many medical practitioners, the cost-reducing effects of Telemedicine are negligible or even non-existent. Cost savings in travel time tend to be only important for medical practitioners in rural and underpopulated areas. Also, the patient's costs of travel are borne by the patients and no cost savings accrue to the doctors. In fact, Telemedicine may even have a negative economic impact for some doctors by disrupting referral patterns and eliminating some sources of income (Abt and Associates).

Social Advantages:

Telemedicine has the potential of reducing of the isolation of medical professionals and offers some social advantages in the form of new, and potentially more satisfactory, interaction among people in the medical field. CTM's survey of the participants of Global Telemedicine 2000 conference in June 1996[2] substantiates some of the economic and social advantages that Telemedicine affords. For example, as shown in

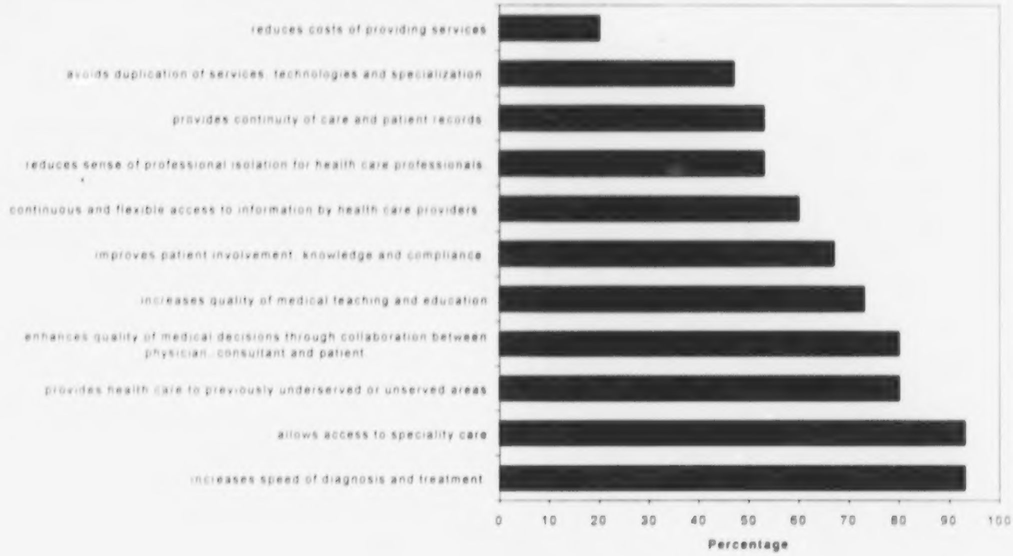
---

[2] It should be noted that CTM's survey of participants at the Global Telemedicine Conference was not a random sample but a survey of 100 specialized medical professionals who were already using telemedicine or were in the process of establishing telemedicine projects.

BEST COPY AVAILABLE

Figure 1, over 75% of participants who have had 20 telemedicine consultations or more a month have observed enhanced quality of medical decisions through collaboration, provision of health care to previously underserved or unserved areas, access to speciality care and increased speed of diagnosis and treatment to some or to a great extent. However, it is equally important to note that only about 20% of respondents had observed Telemedicine leading to a reduction in costs in providing services. This may reflect the high overhead costs currently related with Telemedicine projects, ranging anywhere from $50,000 to $100,000 to equip a typical interactive video site (Perednia and Allen, 1995). This represent a significant barrier to Telemedicine.

FIGURE 1
Percentage of Respondents with 20 Telemedicine Consultations or More per Month Who Have Observed the Following Benefits to Some or Great Extent



Reduction of Uncertainty

Telemedicine also requires sophisticated hardware and high bandwidth[3] as most Telemedicine applications need to be real-time, and "the more challenging and difficult the remote consultation and diagnosis, the higher bandwidth and processing power the clinical application will require" (Kim et. al., 1995). In sum, the technologies supporting Telemedicine are complex and, in a sense, disparate as they need to support videoconferencing, data transfer and database systems. In practice, these separate components must perform as an integrated unit to the user, hence accentuating the importance of user interfaces and information exchange standards. In general, however, Telemedicine can be characterized as involving a high degree of uncertainty.

Compatibility:

Although medicine is an information-intensive professions (Lindberg, 1994) and "every medical encounter is also an information transaction" (Burgener and Kienz.), the compatibility of Telemedicine with current practices and values is low, since there is a long tradition of personal contact between doctor and patient. For example, in 1990, an AMA survey showed that 85% of those surveyed were "very satisfied" with their last visit to a doctor and 90% were "pleased" with the way they had been treated (Wasley, 1992). This

---

[3] It should be noted that Telemedicine applications can be implemented over the Plain Old Telephone System, like the six projects "emphasizing telephone-related technologies: phone, fax, slow scan video, audiographics" reviewed by Witherspon et alli (93).

BEST COPY AVAILABLE

lack of a tradition of instrument-mediated contact between patient and doctor is a major obstacle and the replacement of human, personal contact (high touch) by machine intervention (high-tech) might require a change in the present culture of medicine. Although compatibility is higher in certain medical specializations, where the contact between the MD and the patient is mediated by equipment, like Radiology, early experiments with Telemedicine in these areas failed for technical reasons. Also, numerous Telemedicine applications require large bandwidth. The quality of a video image depends on bandwidth, so for Telemedicine to become widely used, high-bandwidth must be available.

Complexity

The absence of a technological tradition regarding information technology in medicine negatively affects the perception of Telemedicine's complexity. Some Telemedicine projects have cumbersome user interfaces and require extensive technical knowledge. Studies have shown that user friendliness of equipment is crucial for the success of Telemedicine (Mary Moore, 1995). This sentiment is further reflected in CTM's survey where 93% of participants with 20 Telemedicine consultations or more a month found the use of video-conferencing equipment related to Telemedicine to be very easy or easy to use. Not surprisingly, some 67% of participants found radiology and electocardiogram equipment very easy or easy to use since such equipment have had the longest use in Telemedicine. In marked contrast, however, even among frequent users of Telemedicine, only 53% percent of participants found imaging retrieval systems very easy or easy to use, while thirty-three percent found these systems to be very difficult or to somewhat difficult to use. Similarly, some 40% of participants found integrating patient records very difficult or somewhat difficult to use and only 30% found these systems to be very easy or easy to use, reflecting, in part, the incompatibility of the systems.

Furthermore, the lack of standards in Telemedicine hardware, software and networks limits not only modular upgrading of the technological base, but also increases the cost of improvements. For example, although the DICOM standard was adopted in 1985 as the common format for digital medical imaging systems and several different vendors claim that their equipment conforms to that standard, many practitioners of telemedicine assert that images are not transparently interchangeable between vendors (Frederick George III, MD, 1996).[4] CTM's survey substantiated the importance of common standards, training of physicians in the use of Telemedicine, user friendliness of equipment and image quality. Over 80% of participants with 20 or more Telemedicine consultations per month rated the implementation of standards and specifications for procedures, equipment, personnel, licensing and quality control as important or very important factors in Telemedicine.

Trialability

The ability to experiment with telemedicine services on a limited basis before adoption is low, since there is a requirement for specialized equipment and infrastructure. In Telemedicine, particularly, the absence of technical standards and "off-the-shelf" solutions makes trialability even more limited.

Observability

One of the greatest barriers to the increased use of Telemedicine, currently, is the lack of observability of its benefits, since the benefits are usually limited to the participants of the network with a small spill-over effect. CTM's survey found that only 26% of participants within a major city and only 7% of those in a rural county or small town observed or experienced Telemedicine leading to a some or great reduction in the costs of providing services. Similarly, only 27% of participants within a major city and only 13% of participants from rural counties or small towns have observed Telemedicine avoiding duplications of services, technologies and specialization. Furthermore, only 47% of participants from rural counties or small towns have observed Telemedicine providing health care to previously underserved or unserved areas as compared to 61% for participants within a major city.

Overall, CTM's survey found that Telemedicine affords two major advantages for rural communities or small towns, namely allowing access to specialty care and providing continuous and flexible access to information by health care providers: some 53% of participants in rural counties or small towns have observed

---

[4] Personal interview with Dr Frederick George, III in April 1996 at USC Health Sciences Campus.

Telemedicine providing continuous and flexible access to information by health care providers as compared to 37% of participants within a major city.

Social System

Although, the social system surrounding the adoption of Telemedicine is very structured and complex, the lack of a clear position on Telemedicine, in general, by the American Medical Association and most medical colleges and medical schools, save the American College of Radiology, presents another impediment. This lack of clear positioning and ambivalence has contributed, in part, to four major social impediments to the increased use of Telemedicine.

In the first instance, the cost of implementing a telemedicine infrastructure is a large impediment to widespread use of the technology. CTM's survey results confirm this. Currently, a large majority of Telemedicine initiatives are sponsored by organizations where reimbursement is not crucial, like research centers, the Armed Forces or State-owned hospitals, since these are frequently financed by demonstration grants. Only an extremely small number of for-profit medical centers are involved in Telemedicine and many of these, like the Mayo Clinic, are employing closed Telemedicine systems (Tangalos, 1994). Furthermore, medical organizations are reluctant to purchase equipment because of the risk that it will be quickly outdated.

New legislation shows promise in overcoming the payment issue of telemedicine. In California, for example, legislation prohibits state payers from making face-to-face contact between physician and patient a condition of payment. On the federal level, President Clinton has signed a bill that requires reimbursement for telemedicine in rural areas. Payment still does not include reimbursement for telephone line charges or facility fees.[5] However, this is a positive step forward that could pave the way for expanded reimbursement for telemedicine services.[6] Until now, Medicare routinely paid only for radiologists to read images via store-and-forward telemedicine.

Secondly, under the present individual state licensure system the potential of Telemedicine is limited to the somewhat arbitrary borders of a state, thus limiting geographic reach. A new system, enabling physicians to take full advantage of communication networks, should be implemented in order to unleash the potential of Telemedicine.

Thirdly, "there is significant uncertainty regarding whether malpractice insurance policies cover services provided by Telemedicine" (Western Governor's Association, 1995). The legal problems associated with Telemedicine malpractice liability are especially intricate when services crosses state borders. Liability is a significant problem for doctors as shown in a survey by the *Washingtonian* magazine which concluded that seventy-eight percent of physicians are engaged in practicing "defensive medicine"[7] with the result that malpractice liability premiums increased at an average annual rate of some twenty-two percent during the 1980s (Wasley, 92).

Finally, like other communications technologies, there is a concern regarding the security of personal medical information stored in Telemedicine systems. Sanders (94) notes the possible use of encrypting algorithms and legal precedent (yet to be defined) determining "reasonable and customary" efforts in protecting individual's information.

The importance of these issues is substantiated in CTM's survey, where, as shown in Figure 2, over 70% of the respondents with 20 telemedicine consultations or more a month viewed the lack of a universal system of reimbursement as a serious or very serious barrier to the increased use of Telemedicine. In addition, over 50% of the respondents viewed the lack of standards and the incompatibility of state laws as serious or very serious barriers.

Type of Innovation-Decision

---

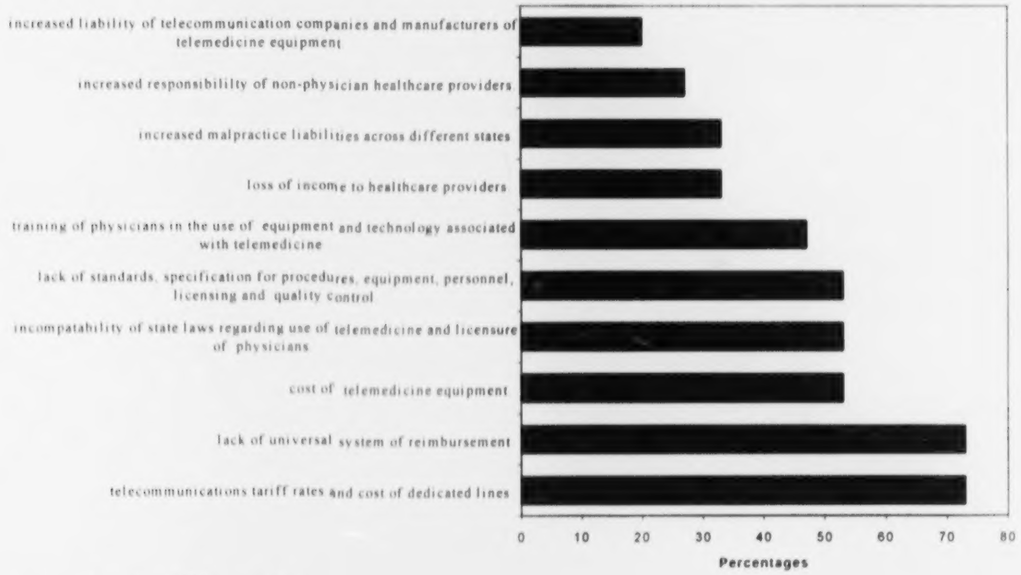
[5] "Congress Issues a Medicare Telemedicine Payment Mandate," Health Data Network News, August 6, 1997.

[6] Under the Budget Reconciliation Act of 1997, Medicare will pay for teleconsultations involving a beneficiary residing in a county in a rural area designated as a "health professional shortage area." About 3.3 million Medicare beneficiaries live in the affected rural areas. Estimates from the Congressional Budget Office show that reimbursement will cost $200 million during the first five years, offset by savings of about $50 million.

[7] recommending possibly redundant or unnecessary procedures only to reduce the risk of malpractice suits

The decision of implementing most current Telemedicine projects seems to be authority based, where users (especially doctors) are not participant decision-makers. Most projects are initiated by policy-makers, like State Public Health officers and Armed Forces leaders. In the future, one can expect a move to a consensual decision-making process for adoption. Several case studies of Telemedicine projects have shown that the success of many of these projects can be attributed to the organizational culture, commitment of management to adopt Telemedicine and the administrative efficiency of the organizations (Moore, 1993).

FIGURE 2
Percentage of Respondents with 20 Telemedicine Consultations or More A Month Who Rate the Following as Very Serious (5) or Serious Barriers (4) to the Increased Use of Telemedicine



Communication Channels

The interpersonal communication among physician and health care administrators is a primary source of communication regarding Telemedicine decisions. Other sources include vendors of equipment and services[8], the professional journals and the media. Interviews with telemedicine directors have found that those leading such projects tended to be charistmatic entrepreneurs, articulate, enthusiastic, energetic, self-sacrificing, obsessed with their users, impatient for change and true believers in their cause. Furthermore, physicians who were most likely to use telemedicine were described as being inquisitive, confident, demonstrating qualities of lifelong learning, preferring to use many sources for information, and were often outgoing (Williams and Moore, 1995). A communication barrier for Telemedicine is the fact that a high proportion of adopters are the small rural hospitals. Doctors and administrators from rural areas are not especially well positioned to serve as a reference group for medical professionals and institutions overall.

Conclusion

At a macro level, the diffusion of Telemedicine is being accelerated by a concern with health care costs and demographic changes. The cost pressures of health care have already forced major changes in the sector structure; the emergence of the Health Maintenance Organizations (HMOs), non-existent in 1970 and now with more than 56 million beneficiaries, probably best exemplifies this. The demographic changes, specifically the aging of the population in the U.S. and most industrialized countries, are generating social pressures in favor of the higher productivity that Telemedicine can bring (Gott ,1995).

---

[8] Pacific Bell, for example, sponsors the Telemedicine project at the University of Southern California.

BEST COPY AVAILABLE

The eventual large-scale adoption of Telemedicine could cause radical changes in the structure of power and interests in the medical profession, in particular, and society, in general. These potential outcomes may further act as a barrier against its wide-scale adoption. In the first instance, the massive adoption of Telemedicine would certainly require a very different organizational arrangement, bringing substantial changes in the way medicine is practiced. The full effectiveness of Telemedicine, however, will only be achieved when some medical responsibilities are delegated to physicians' assistants and nurse practitioners. This could lead to a power transfer to these groups, with considerable modifications in the differential social standing of doctors and other medical personnel.

Secondly, the legal and operating restrictions on the practice of Medicine have protected the medical profession against intense competition and created a near-oligopoly in the heatlh care industry. Telemedicine has the potential of reducing the barriers to competition, giving patients more treatment options and increasing competition among health care providers. Furthermore, Telemedicine will not only enable competition among doctors of different states or even countries, but also between medical doctors and other medical personnel, like nurse practitioners now empowered by Telemedicine to treat cases previously referred to a general practitioner.

In the final analysis, however, the full potential of Telemedicine will only be realized through: i) change in medical culture and attitudes; ii) changes in the model of health care delivery; iii) current methods of funding requirements from state and federal sources restrict commerical opportunities for equipment leasing and data storage iv) cooperation and coordination between corporations, government bodies and health care providers; and v) definite analyses of the costs and benefits, both economic and social, for Telemedicine.

REFERENCES

Abt Associates Inc. 1995 Rural Telemedicine Evaluation. http://www.arentfox.com/ telemed.rural.html.

Burgener, Alan J. and Michael G. Kienzle. A Test Drive On The Health Care Information Superhighway. http://indy.radiology.uiowa.edu/Welcome/Sponsors/NLM/BergernerKienzle.html

Clyburn, Conrad A. An Integrated Process Re-engineering Model for Improving Health Care Delivery Using Telemedicine. World-wide web document. http://www.matmo.army.mil/pages/library/papers/reengineer/procreng.html

Gott, Marjorie. Telehealth and Telemedicine, 1994. Executive Summary of the European Foundation for the Improvement of Living and Working Conditions Research Project. Luxembourg: Office for Official Publications of the European Communities.

Kim Y., J. E. Cabral Jr., D. M. Parsons, G. L. Lipski, R. H. Kirchdoerfer, A. Sado, G. N. bender, F. Goeringer, 1995. SeaHawk: A Telemedicine Project in The Pacific Northwest. http://icsl.ee.washington.edu/projects/gsp9/spie95/seahawk

Kim, Donlok, James E. Cabral Jr. and Yongmin Kim. 1995. Networking requirements and the role of multimedia systems in Telemedicine. http://icsl.ee.washington.edu/projects/gsp9/spie1095.

Lindberg, Donald A. B. Statement. 1994. Written testimony to the Telemedicine hearing before the Subcommittee on Investigations and Oversight, Committee on Science, Space and Technology, U.S. House of Representatives, 103th Congress. U.S. Goverment Printing Office.

Moore, Mary, 1995a Elements of Success in Telemedicine Projects. http://naftalab.bus.utexas.edu/nafta-7/elemsucc.html.

Moore, Mary 1995b. Telehealth Cost Justification. http://naftalab.bus.utexas.edu/nafta-7/costjust.html.

National Telecommunication and Information Administration, 1991. The NTIA Infrastructure Report: Telecommunications in the Age of Information. Special Publication 91-26. U.S. Department of Commerce.

Office of Technology Assessment, 1990. Health Care in Rural America, OTA-H-434. Government Printing Office, Washington, DC.

Perednia, D. and A. Allen, 1995. Telemedicine technology and clinical applications Journal of American Medical Association, 483-488

Rettig, R. "Medical innovations duels costs containment" Health Affairs, 13, 7-27, 1994.

Rogers, Everett M. 1995 Diffusion of innovations. New York: Free Press, 4th ed.

Sanders, Jay H. Telemedicine, 1994 : Challenges to Implementation. Written testimony to the Telemedicine hearing before the Subcomittee on Investigations and Oversight, Committee on Science, Space and Technology, U.S. House of Representatives, 103th Congress. U.S. Goverment Printing Office.

Schwartz, M., 1994 Health Premium Growth Rate Continues to Slow, National Underwriter Life & Health-Financial Services. 43, 8-10.

Tangalos, Erig G, 1994 . Telemedicine: an Information Highway to Save Lives. Written testimony to the Telemedicine hearing before the Subcomittee on Investigations and Oversight, Committee on Science, Space and Technology, U.S. House of Representatives, 103th Congress. U.S. Goverment Printing Office,

Wasley, Terree P. 1992. What has Government Done to our Health Care? Washington, DC: Cato Institute.

Western Governors' Association. Telemedicine Action Report. http://www.arentfox.com/telemed.western.html

Williams, Frederick and Mary Moore. Telemedicine: Its Place on the Information Highway. http://naftalab.bus.utexas.edu/nafta-7/telepap.html

Williams, S and P. Torrens, 1993 . Introduction to Health Sciences Albany, Delmar.

Witherson, John P., Sally M. Johnstone and Cathy J. Wassem. 1993. Rural Telehealth: Telemedicine, Distance Education and Informatics for Rural Health Care. Office of Rural Health Policy, U.S. Departament of Public Health and Human Services

WEB-BASED EDUCATIONAL MEDIA: ISSUES AND EMPIRICAL TEST OF LEARNING

Senthil Radhakrishnan
James E. Bailey, PhD
Industrial and Management Systems Engineering
Computer Integrated Manufacturing System Research Center
Arizona State University
Tempe, Arizona, 85287-5906
Phone (602) 965-5346
Fax (602 965 2910
(E-mail) senthilr@asu.edu
J.Bailey@asu.edu
www.eas.asu.edu/~ece300/

Abstract: Web-based educational media are being developed rapidly and the pressure to employ this technology for distance learning is growing. Educators are rightfully asking questions regarding the cost/benefit of such efforts and how authors might deal with expected problems when employing such an open media. More basic research into these questions is needed. This paper attempts to shed some light in this regard. The results of an effort to identify problems that arise when considering this media are presented. Conceptual solutions to some of these problems are suggested. To test the concept, a prototype system was built and tested in an engineering classroom and the educational results of that test are presented. The research effort reported here was, in part, funded by the National Science Foundation.

Significance

Paper presents concept and empirical results that can contribute to the improved design of Web-based distance learning media.

1. INTRODUCTION

The low operating requirement and great potential audience for Web-based educational delivery has generated great interest in this technology. A few years ago, one could find only experimental courses typically built by computer science faculty. Today courses at all levels of education in many fields are being reported. In the Western U.S., an initiative to create an "open university" wherein students from several Rocky Mountain States will be able to attend courses at several universities has been established by those state's respective governors. Web-based distance learning systems are a big technological wave that is fast approaching.

Yet there are many unanswered questions regarding Web-based educational systems. The primary question is, whether the student can learn as well using this technology as when taking courses in the traditional classroom. A highly related question is the potential number of added student that can be reached through distance learning. A third question is the relative magnitude of development and maintenance costs as well as delivery costs when using such course systems. Technical questions arise regarding the protection of intellectual property rights including the broadcasting of copyrighted materials. Answers to these and other questions are needed as these system begin the move into the mainstream of education.

This paper is one small attempt to contribute to the body of information regarding the issues and cost/benefits of Web-based education systems. The results of an effort to identify problems that arise when considering this media are presented. Conceptual solutions to some of these problems are suggested. To test these solutions, a prototype system was built and tested in an engineering classroom and the educational results of that test are presented.

2. ISSUES WITH WEB-BASED EDUCATIONAL MEDIA

There are a number of issues involved with web based educational media. Some of them are hurdles or delimiters in this new concept, whereas there are others, which encourage usage of this media to transport education. The next few paragraphs describe the issues involved with developing and using such a system.

2.1 Favorable aspects of web based educational systems:

These are a growing number of web authoring tools which help in developing courseware for the web. The advantage is that the user need not know anything about HTML or other programming languages, and the entire task of placing the courseware on the web in an orderly fashion is accomplished by the tool itself. An example of an authoring tool is HCC(HTML Course Creator) which allows instructors who are not HTML experts to rapidly develop and easily maintain consistent libraries. The course can be tailored to specify styles based on templates to cater to various types of courses. Using tools like HCC instructors from different backgrounds can create and maintain network hypermedia courses accessible over the web.[Carver 1996].

Research is going on to address the security issues in the sale and distribution of information over the Internet. IBM infomarket is a new network based service offering from IBM that will allow digital publishers to sell their content over the web. This service will support the secure distribution of intellectual property to limitless number of "downstream" consumers over electronic networks, while providing a mechanism for the copyright holder to receive payment for each use of the subject matter. This technology will encourage more and more authors to have their writings on the web.[Crigler 1996]

Another crucial advantage of web based systems is the role they play in long distance education. Since the course is on the web, it can have a very large audience. This also means that the system can be made cost efficient and there is scope for improvement. With multimedia technologies developing so rapidly a multimedia approach requires a whole new approach to the learning process. Multimedia, e-mail and online quizzes all packed together in the system revolutionizes the whole concept of Distance Education giving it a new definition.

Having a course on the web, directly places a student in a computer environment. This give the student direct access to many other software tools like search engines, word processors and spreadsheets which are often required to do assignments. In this way it offers many indirect benefits.

2.2 Problematic aspects of these educational systems:

The web based educational system has to be evaluated before any large-scale implementation is carried out. A good evaluation of the effectiveness of this system will come as more and more faculty make use of it and report their experiences to their peers and administrators. In one particular case the faculty member offered two sections of the course one taught through traditional lecture means and the other taught by using the WWW in all facets of course work. In the traditional set-up students were distributed materials or pointed to libraries, they wrote their reports with word processors and gave in reviews of their research. In the web based section students were directed to readings on-line and library resources. All student research and reports were put together and delivered on the web. Students using the web-based system were also able to collaborate partially. For both classes the average grade was 'B', but most effective feedback came from the students themselves. In the web based class students felt they invested more time in projects, had a steeper learning curve, the collaborative process was fruitful and the sense of accomplishment was greater.[Ellen et al. 1995].

Developing an entire course is still an expensive affair. Authoring tools are still immature and often too generic. The HTML converter tools for example are not very efficient. Thus the instructor is forced to learn more about the web. This may discourage the faculty in developing their courses for the web. Also putting courses on the web is a long process, especially if one wants to make it interactive with a lot of multimedia features. One also needs an expensive digitizer to convert analog video signals to digital format to store it in a disk. But these issues are being addressed and one can expect more tools in the future, which are less expensive.

Copyright issues of documents placed on the web are not settled. The security features that are available will not prevent a student from making multiple copies of a document and distributing it. This aspect of the web will discourage many authors from putting their works on the web. Employing tools such as IBM's infomarket, some solutions are available, but questions on how one could incorporate these features are still unanswered.

When using the web, one very important issue is that the presentation of the course material should be far superior to how it is currently done with overhead projectors or for that matter with textbooks. The web-based system should be highly interactive and should incorporate lot of multimedia features, which can give the learning environment a new and refreshing flavor.

The Internet is still a very slow communication medium. Internet access by students outside the university campus is typically frustrating because the current data transfer rates are very low. Also the processor speed on the client machine makes quite a difference. A slow processor can greatly increase the time to display images. Therefore the web is still not suitable for big files with many images or graphics. However, it is suitable for text materials.

Today, students do not have sufficient access to computers and many students have no computer at home. This factor can web based education more expensive for students and is very discouraging in developing web-based courses. When one is dependent on someone or something else it makes learning a less interesting process. The cost of a computer with the required configuration is quite high.

3. CONCEPTUAL DESIGN OF A WEB-BASED EDUCATIONAL SYSTEM

The solutions to problems suggested in section-2 will require technical concepts similar to those developed in an National Science Foundation funded project to connect product design teams to each other and wide varieties of proprietary design data [Bailey]. In this section we shall describe a conceptual design and first prototype of a system for Web-based education which offers solutions. Much work is needed for such a system to be completely available in its totality.

3.1 Functional Requirements

- Easy and low cost transfer of existing course materials.
- Easy and low cost development of new materials.
- Timely media-based feedback to students.
- Automatic customization of materials to students.
- Support of various learning styles through multimedia.
- Easy communication to instructors of fellow students.
- Easy links to search engines and course related literature.
- Fast transfer of picture and video material.
- Difficult t to copy or otherwise share material.
- Time management capability to help students schedule.
- Easy access to appropriate learning support tools.

3.2 Conceptual Solution

The objective of the research reported here was to build a web-based system that provides some of the required functionality and then to test the ability of that system to improve the education process a broad conceptual design was designed. The test system was developed with locally stored pictures and videos. The system provided the student with point and click access to word processing and spreadsheet tools.

The system involves the usage of a web server to deliver part of course module over the web with video and graphics supplied from a CD-ROM. When a student in/off campus registers for the course, he/she is provided with a CD-ROM, a manual to use the package and other course material. The client can only view the course from a popular browser like Netscape3.0 or above. Some of the course material is delivered from a web server over the net, whereas most of the video files and other long files are read from the CD-ROM, within the browser environment. The student can also register for the course on-line over the web. The student is then provided with a user_id and password by which he can access the course homepage. The student is also sent a CD-ROM and related material by mail.

All details about the student, quizzes and grades are stored in a database. The text material and any updates or changes in the course are delivered from the web server. The web server maintains the connections to the database server. Having database storage also gives a lot of flexibility in designing the course to cater to a wider group of students. Each student's course profile (the course structure varies from student to student giving him an opportunity to select a course which more closely fits his requirement or interests) is stored in the database, and when the student accesses the course over the web, the course delivered is tailored to that student's background. For example, electrical engineering students would receive different examples than mechanical engineering students.

There is also an IRC or chat server. Students registered for the course can discuss about the course with their classmates. It also allows the instructor to talk to the students once in a while and answer their queries immediately. This feature partially covers the sociability aspects, which is an important element in a typical classroom environment.

The course also has a search engine with predefined bookmarks to get course-related information from the web. Thereby the web server behaves as a online library and a suitable medium to gather more information on the subject. The search engine can also serve as a quick index for the course material.

Basically the system gives a student the facility of being able to sit in one place and do everything, read about the course, take virtual classes, do assignments, discuss the course with his classmates and even answer the tests. One issue not talked about in the system is how the course material is going to be developed and cost analysis of the system.

3.3 Prototype Test System

The prototype system built to test the suggested concepts consisted of two major components: a course-assignment/team-communication sub-system and a lecture-delivery sub-system. These sub-systems were integrated and delivered via Netscape. The system employed a Sun Solaris university server connected to ASU's student network. The reader can access the system via URL: www.eas.asu.edu/~ece300/.

The home page for the system contained HTML buttons or links to access information about the course, instructor, and a real time grade status reporter. The first two pages were open to all users but the last button activated a Java program that access the gradebook database using a PIN number. Thus students had access to only their own grade data. The home page also allowed access to the course-assignment/team-communication and lecture-delivery subsystems.

The course-assignment/team-communication page as illustrated in Exhibit-1 permits any student to highlight any team including his/her own and any team or individual-assignment assigned to his/her PIN number. When a team is highlighted, the members of that team are displayed and the assignments file IDs associated with that team are instantiated. The student can highlight any sub-set of team members and activate an E-mail package. In this way, team members can communicate with each other, the instructor and industrial term project client/mentors. They can also highlight any assignment for which they have PIN-number access authorization and open application software package with the appropriate document. For example, through Microsoft-office, they can access reports in Word or spreadsheet models in Excel. The most recent version copy of the document will come up on their computer. Several problems exist. The student's computer must have the application software on their system. One has to have Netscape3.0 or higher and Windows 95 as the browser has to be Java enabled and should support JavaScript. In addition data concurrency is problematic since more than one

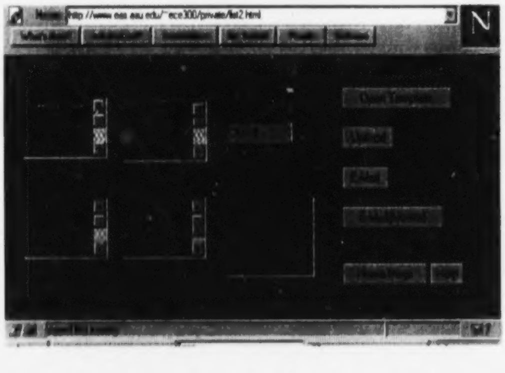
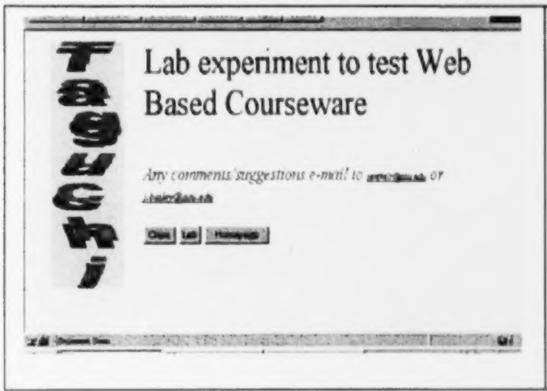


Exhibit 1:
Course assignment
/Team-communication
page

copy of any file can be active. In any case, the student can be given read-only access or read-write access to any output assigned to him/her or to the team.

The lecture homepage is illustrated as Exhibit-2 permits the student to attend any electronic lecture on the system. A list of available lecture topics is given; for the test prototype, only one lecture module was generated. Upon highlighting a lecture the student starts by bringing up a set of detailed lecture notes. These notes are generated to cover what the instructor believes to be a non-reducible concept as illustrated in Exhibit-3. Present experience suggests that a typical 50-minute lecture would translate into between 5 and 10 concepts. The level of detail, in these notes, is sufficient for a student, familiar with the material, to review the issue.



Lab experiment to test Web Based Courseware

Any comments/suggestions e-mail to [illegible] or [illegible]

Exhibit 2: Course homepage

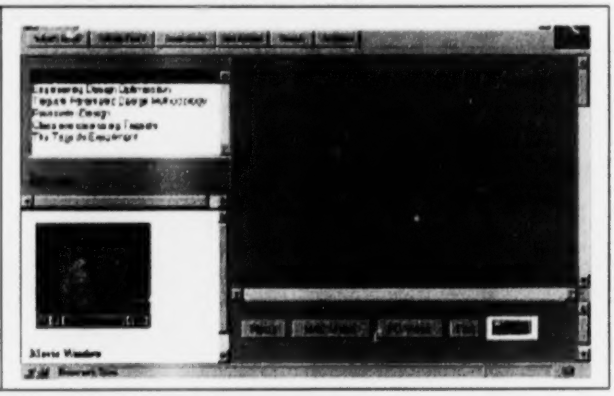


Exhibit 3: Course Presentation

Three HTML buttons are available for each issue: a text button, a video-lecture button, and a quiz button. Clicking the text button delivers a window containing a textbook level written discussion about the issue. These notes reside on the CD-ROM that the student purchased so the material is viewed from a web browser only. However, while the student's PIN number is active, he/she may access the text material from the web server as often as desired. The video-lecture button to the student delivers a videotaped lecture stored in the CD-ROM. Here too, the video is difficult to copy but can be viewed often. Finally, a quiz button when clicked delivers a computer graded practice quiz to the screen. A random subset of computer gradeable questions is printed allowing the student to test his understanding. Upon completion, the computer will grade the quiz and announce the score. The student can than decide if they need to spend more time on the topic or go to the next module. All these sub-pages have buttons to return to the lecture note.

4. PILOT TEST OF A WEB BASED EDUCATIONAL SYSTEM

Conducting such a sophisticated study that compares two different sorts of course environments to find if one is "better" than the other is practically impossible. What need's to be done is specify certain conditions and compare aspects of one environment to parallel aspects of the other environment. But either so, one will not get a simple-to-interpret answer of "is a better than b?"[Collis 1997] If the test is conducted in the real world, due to gross differences in the types of courses and other environmental factors, even a reasonably accurate assessment of the system is doubtful. To avoid some of these problems the test is conducted in a lab environment with preset parameters, and any discrepancies in evaluation due heterogeneity of the students in the system can be washed away by statistical techniques. But the only drawback with such a approach is that students tend to be conscious of being part of an experiment and may tend to behave in a less natural manner, which may affect the final result. A lab based approach to evaluate the educational system is described below.

The prototype system described above was used to test the knowledge transfer and attitude of students as compared to the traditional classroom system. Two hypotheses were tested:

Ho(1): The amount of learning via the Web-based system is no better than that of the traditional face-to-face delivery system.

Ho(2): The students attitudinal reaction to the Web-based system is no better than toward the traditional face-to-face delivery mechanism.

BEST COPY AVAILABLE

These two hypotheses were to be tested empirically and if both are rejected, we could conclude that a Web-based educational delivery system was shown in this case to be superior to the traditional system. Because these results are not available as this paper is written but will be available before the paper is presented, they are not presented in the paper but will be included in the Toronto meeting.

The test scenario was an one week engineering course module covering the Taguchi method for establishing robust design parameters. The text, notes, and lecture materials were adapted and delivered by the research advisor co-author of this paper. Two classes of 35 students in Arizona State University's Engineering ECE-300 were used as test subjects. The test was run in the same week of April 1997. The module consisted of one 50-minute lecture covering the Taguchi methodology, a 50 minute laboratory exercise in which the student collect data and analyzed the parameters of a toy catapult, a 15 minute quiz taken manually by both sets of students, and an attitude questionnaire. Attitude was measured using the semantic differential technique (x)[Bailey 1983]. The test data was, therefore, the quiz scores and the attitude scores.

Demographics for the students were collected and scores adjusted to account for potentially affective differences in the two test populations. The student's ages were assessed as a surrogate for maturity, which might affect attitude and performance. Class standing data, in terms of credit hours taken as a surrogate for the added experience more advanced students would bring to the test. Three times the number of credit hours being taken plus the hours being worked per week were measured as a surrogate for the time a student would have available for study. Finally, the students' grade point average was collected as a surrogate for the students' intelligence and seriousness toward school.

The primary conclusion was that the Web based system resulted in significantly better learning as measured by an average of 10 more points on a 75 point quiz. The greatest increase was for questions that came from instructor provided class notes. The study was unable to reject the hypothesis that the students reaction to the experience was any different between the two systems except that students preferred the hands-on laboratory exercise over its computer simulation. None of the demographic factors: age, GPA, or credit hours affected these conclusions. Finally, students working with the Web based system spent more time studying the subject, which did affect their performance. However a regression analysis showed that students using the traditional method improved their grade more per hours of study (2.18 to .75 points/hour) than did students using the web method.

5. CONCLUSION

The paper began by giving an overall picture in implementing a Web based educational media. The various issues in moving towards a web based educational media were also discussed. A list of requirements for a web-based system was enlisted and a conceptual system was proposed which aims at satisfying most of these. A prototype system developed and implemented was discussed. Finally, an evaluation methodology for a system just developed and the results of the experiment was given comparing the web-based system and traditional classroom

The whole world is moving in the direction of computers and Internet, and we are in the threshold of an Information Revolution, unparalleled even with the advent of the television and the telephone. Sooner or later like many other fields in life, even Educational styles have to shed some of there traditions and jump into the Internet bandwagon if it has to keep pace with changing technologies and lifestyles. With current technology limitations matters don't seem to be all that favorable for a web based system but it is definitely going be the solution in the near future. Does this mark the doom of the traditional classroom?

REFERENCES:

[Carver 1996] Carver, Clark Ray- Automating Hypermedia Course Creation and Maintenance: pg82 Proceedings of WebNet'96

[Crigler 1996]Crigler-Offering Services on the Web: pg121 Proceedings of WebNet'96.

[Ellen et al. 1995] Ellen, David, Larsen, Deborah- Supporting teaching and learning Via the Web: Transforming Hard-Copy Linear Mindsets into Web-Flexible Creative Thinking - pg44 Proceedings of WebNet'96.

[Bailey] Bailey and Rucker. "Automated Design Data Management Product, Process and Resource Structures, "(accepted by International Journal of Industrial Engineering.

[Collis 1997] Betty Collis-University of Twente by E-Mail.

[Bailey 1983] Bailey and Pearson, "Development of a Tool for Measuring and Analyzing Computer User Satisfaction," Management Science, Vol. 29, No.5, May 1983.

Visualization in a Mobile WWW Environment

Alberto B. Raposo
Dept. of Computer Engineering and Industrial Automation (DCA)
School of Electrical and Computer Engineering (FEEC)
State University of Campinas (UNICAMP) - Campinas, SP, Brazil
E-mail: alberto@dca.fee.unicamp.br

Luc Neumann
Dept. Mobile Information Visualization
Computer Graphics Center (ZGDV) - Darmstadt, Germany
E-mail: neumann@zgdv.de

Léo P. Magalhães[1, 2]
[1] Computer Science Dept. - Computer Graphics Lab.
University of Waterloo - Waterloo, Ontario, Canada
[2] DCA - FEEC - UNICAMP - Brazil
E-mails: lpini@cgl.uwaterloo.ca, leopini@dca.fee.unicamp.br

Ivan L. M. Ricarte
DCA - FEEC - UNICAMP - Brazil
E-mail: ricarte@dca.fee.unicamp.br

Abstract: The facility of access to information in the World-Wide Web (WWW), the expanding availability of information technology, and the recent developments in the handling of multimedia data are all important steps towards a Global Information Infrastructure (GII) accessible to anyone, anywhere in the world. However, in order to achieve this accessible infrastructure, one should consider the aspects related to efficient communication. These aspects are addressed by this paper. A mobile WWW rendering application using VRML 2.0 (Virtual Reality Modeling Language) is introduced, the related problems are pinpointed and approaches to overcome them are proposed. We have developed an application that filters VRML scenes to render only parts selected by the user. It improves the interactive visualization within a mobile environment and is a further small step towards the GII.

Introduction

The emergence of the WWW expanded the global information space, integrating services offered by the Internet and other networks. However, the number of users is increasing rapidly and the information they need is becoming more complex, requiring new ways of communication and interaction. The challenge is to reach the Global Information Infrastructure (GII) [Gershon et al. 96], a step beyond the WWW with more efficient interaction and globally accessible.

The technology of mobile communications is becoming increasingly integrated into the WWW. Due to this rapidly expanding technology, mobile users with portable computers are capable of accessing information anywhere and at anytime.

In general, a mobile WWW application uses mobile devices such as Personal Digital Assistants (PDAs) or notebooks to access remote WWW services. The required connection to the stationary WWW server can be established with communication facilities provided by the mobile device itself or with a cellular phone. The scenario is illustrated in [Fig. 1].

Title: Mobile_static.PS
Creator: XV Version 3.00 Rev: 3/30/93 - by John Bradley
CreationDate:

Figure 1: Mobile Application Scenario

This new communication style significantly affects the requirements of WWW applications and faces new challenges [Forman & Zahorjan 94], [Satyanarayanan 96]. One challenge is the design and implementation of highly interactive applications handling non-textual bulky information.

The main problems in handling such applications in a mobile environment are the resource constraints of the mobile data terminal and the narrow bandwidth of wireless wide-area networks. Thus, sophisticated concepts for the management and optimal utilization of resources within a mobile system are needed.

In this paper we address the remote access of VRML 2.0 worlds [VRML 97] from a mobile client, typically a laptop-like environment. In Section [Challenges of Mobile Computing] we present the peculiarities of mobile computing and its challenges. Our solutions are introduced in [Application Scenario and Solutions]. In [Conclusions and Future Work] we come to the conclusions and propose future improvements on this work.

Challenges of Mobile Computing

A mobile environment is characterized by at least the following properties [Neumann et al. 96]:

- Limited resources of the mobile devices in terms of storage, battery power, memory and processing power relative to non-portable devices.
- Low bandwidth, low reliability and high costs of wireless narrowband wide-area networks (WANs). Typical bandwidths of currently available cellular systems are 9.6 Kbps (GSM - Global System for Mobile Communication) or 19.2 Kbps (CDPD - Cellular Digital Packet Data), insufficient for multimedia applications.
- Imbalance of resource availability between the mobile device and the stationary servers.
- Users have no fixed location and can move during a connection.

The introduction of interactive applications - such as a WWW application - used within a mobile environment will only be successful when the requirements arising from the handling of non-textual bulky data (e.g., graphics, animation) can be fulfilled.

A simple solution might be to apply all the well-known mechanisms for a distributed application in a wired network also for a mobile application. However, they are designed for a higher bandwidth and richer resources at the end device. It will work, but the requirements in terms of throughput, delay, jitter, or response time will not be fulfilled. Therefore, it is a challenge to develop appropriate techniques such as compression, progressive refinement, and previewing, for a mobile environment to solve these problems.

In general, solutions address two main aspects related to the specific properties or limitations of a mobile environment.

- The transferred amount of data has to be as small as possible. This requires that parts of the application data be processed and stored on the client side.
- The use of local resources such as processing power and storage space should be the least possible. The client provides a presentation front end and communicates frequently with the server that processes computing intensive parts of the application.

It can be easily seen that solutions for one aspect are counterproductive for the other aspect. A solution providing an appropriate trade-off between both aspects would overcome the most serious problems - namely the narrow bandwidth and limited resources - of a mobile environment.

Application Scenario and Solutions

We focus on a rendering scenario where mobile users connected to the Internet via a wireless communication channel request for the visualization of a certain VRML world. The description of the world is stored somewhere on an information server. This description will be retrieved, rendered, and presented on a mobile client. However, the narrow bandwidth and the restricted processing power of the mobile data terminal require intelligent strategies to enable an interactive handling of the rendered scenes. Simply stated, the straightforward approach of retrieving the scene description and rendering it on a mobile client is not an adequate solution, especially if we think of complex scenes.

The solution we adopted in our application is a kind of data filtering technique. In a further improvement of the application, we propose the use of server resources in the rendering process, which will decrease the resource utilization in the client side. This enables a good balance between bandwidth and resource requirements in the mobile data terminal and ensures a good quality of service.

Reducing the Scene Complexity

Rendering is notably a process that requires a large amount of processing power. Due to the limited resources of mobile devices, techniques are necessary to simplify that process. One straightforward approach is to filter the data to be transmitted and to render only the parts of the scene that are actually of interest.

In order to demonstrate this strategy, we have developed an application capable of selecting the elements (i.e., geometric objects, light sources, and cameras) of a remote VRML world, which are going to be rendered and visualized in the mobile client [Raposo et al. 97]. This application runs in a mobile terminal (as a Java applet [Campione & Walrath 97], downloaded via an HTML page), which is connected to an application server. This server is capable of reading the VRML world (located in any WWW-server), parsing it and sending its hierarchical structure to the client. The client provides a user interface adapted to the hierarchical structure, enabling the users to select the elements of the world they want to see. The selected elements are then sent to the application server, which can parse the original VRML file and extract from it only the desired elements. This valid "sub-VRML" world is sent to the client, which can finally render the scene. This approach is illustrated in [Fig.2].

Title: cl_serv.eps
Creator: fig2dev Version 3.2 Patchlevel 0-beta2
CreationDate: Tue Jul 1 13:31:19 1997

Figure 2: Strategy to visualize only a subset of a VRML 2.0 world.

The process begins by calling the application using any Java-enabled WWW browser. At this moment, a connection is established between the client and the application server. The next step is to send the URL of a VRML world to the application server (arrow 1 in [Fig. 2]). The application server will then connect with the server of the VRML file, read and parse it (arrow 2 in [Fig. 2]). The first output of the application server is a small document (called *scene graph document*), representing the hierarchy of the elements of the VRML world, which will be sent to the client (arrow 3).

Based on the received document, the applet creates an interface, which allows the users to select the desired elements of the world. After this selection, a new document is sent to the application server, describing the selected elements (arrow 4 in [Fig. 2]). Using this new document, the application server can create a new VRML world from the original one, by extracting only the desired parts of it. This final sub-VRML world is then sent to the client (arrow 5), that visualizes the results using any VRML browser connected to the Web browser (as a plug-in or helper application).

Object Selection

Although only textual data are transmitted when the scene is completely rendered in the client, the narrow bandwidth of the wireless connection requires additional data reduction, especially for complex worlds. Our solution asks the user to build a sub-VRML world to be sent to the client. This sub-VRML file becomes smaller if the user selects fewer elements to be visualized. By presenting only the elements selected, which simplifies the rendering process, the utilization of the resources on the client side is reduced.

The application has also the interesting capability of mixing elements of different VRML worlds. In this way, the user can read several worlds, selecting and combining elements from each of them.

[Fig. 3] shows the interface of our application. In the top of the interface window there is a place for the URL of the VRML world and below, two buttons (*Read it*) allowing to read a new world. The first one (*add*) maintains the previously selected elements in the next visualization, while the second (*new*) removes all elements before loading the new world. The large text area in the middle contains the script used in the transmission between client and server. (This script is showed for demonstration purposes only - the user does not need to deal with it.) Below the script area there are buttons to choose the appropriate elements and to visualize the result of the selection. In the figure, the *Cameras* button has been pressed and the user can select among the several cameras of the world.

Title: application.ps
Creator: XV Version 3.00 Rev: 3/30/93 - by John Bradley
CreationDate:

Figure 3: Interface of the developed application.

Task Distribution

In this section we present an approach to optimize the rendering process of the previous sections by the distribution of the rendering tasks among the available resources of the mobile client and the stationary servers.

This will improve the rendering process within the limits of the available resources. The main approach is to use the knowledge about the application semantic data and the environment resources to distribute the tasks.
The establishment of such an architecture is based on two main aspects: the introduction of a semantic content header for the scene description and the application of Java-based Object Request Brokers (ORBs) to integrate the distribution architecture into the WWW.
For the distribution of the different tasks to the most suitable resources we propose an architecture with a *Resource and Task Manager* (*ResTaMan*) [Neumann et al. 96] that distributes and controls the rendering process using application semantic information. A VRML-analyzer reads the semantic header and the world description. The semantic header can be regarded as metainformation about the scene supporting the identification of subtasks to be distributed. Here, the filter introduced in the previous section may be used to extract the relevant parts of a scene for the different renderers. In combination with an ODP-Trader [ODP Trading 97], which is a yellow page service knowing the properties of the environment resources, a good utilization of the rendering resources on the mobile client and in the fixed network can be achieved. Finally, an image composer and synchronizer is necessary to display the result.
Java ORBs are used for the integration of *ResTaMan* in the WWW environment. This enables the WWW Browser to gain access to arbitrary CORBA-based services. A Java applet is downloaded to the WWW client and serves as CORBA client accessing services using the Internet Inter-ORB Protocol (IIOP) [OMG 95]. In our approach it contacts the *ResTaMan* object that represents a metaserver from the viewpoint of the client. The entire rendering application consists of a set of interworking objects playing together via an ORB. For the transmission of images or image sequences we use a stream-oriented connection between the Applet and *ResTaMan* that may be established with an IP socket connection. Thus, for the bulk data transfer the stream connection is used and the control operations are transmitted via the IIOP. [Fig. 4] illustrates the integration of our architecture with the WWW.

Title: realisation.fig
Creator: fig2dev Version 3.1 Patchlevel 1
CreationDate: Tue Jun 17 10:28:52 1997

Figure 4: Integration of the Distributed Rendering Architecture into WWW.

Because VRML 2.0 also describes interactive worlds, we need a distributed event mechanism to send information about the occurrence of a specific event in the client to the appropriate remote object. CORBA introduced an *Event Notification Service* allowing objects to register their interest for specific events or to inform interested parties about the occurence of events. This service will be used in our environment to handle events and to inform the respective rendering object.

Conclusions and Future Work

There is a growing need for tools supporting the interactive access, manipulation, and visualization of distributed multimedia information to realize the vision of "all information at your fingertip". One great challenge is the improvement of the accessibility in the WWW using mobile devices. The main problems are the limited resources on the mobile device and the narrow bandwidth of the wireless link.

In this paper we presented solutions aiming to provide a good trade-off between bandwidth and resource requirements. We developed an application enabling the user to select elements of a VRML world and proposed an architecture for the adaptive distribution of tasks in a mobile environment. Although our solutions have been developed aiming a mobile environment, they can be used with any other type of WWW client (e.g., dialup).

The main advantage of our approach is the application of interactivity to achieve an efficient transmission. In order to filter the data or to reduce the complexity of the scene (efficient transmission), the user selection is needed (interactivity). In the current prototype, this interactivity still requires some knowledge of VRML by the users, but the goal is to use semantic information (defined by the author in the *semantic content header*) to achieve as much transparency as possible in the process. In this way, the users would be able to select objects by the role they played in the world (as defined by the author). Another possibility is to have the author assigning levels of priority to objects in the world, with a high level to objects which are essential to the scene and a low level of priority to details and textures, for example. In this case, the users would have only to set up to which level of priority they are willing to wait.

References

[Campione & Walrath 97] Campione, M., and Walrath, K. (1997). *The Java Tutorial - Object-Oriented Programming for the Internet.* The Java Series. Addison-Wesley. (http://java.sun.com/nav/read/Tutorial/index.html).

[Forman & Zahorjan 94] Forman, G.H., and Zahorjan, J. (1994). The Challenges of Mobile Computing. *IEEE Computer*, 27(4), 38-47.

[Gershon et al. 96] Gershon, N., Brown, J.R. et al. (1996). Computer Graphics and Visualization in the Global Information Infrastructure (special report). *IEEE Computer Graphics and Applications*, 16(2), 60-75.

[Neumann et al. 96] Neumann, L., Zhang, J., and Strack, R. (1996). Concepts for the Efficient Handling of Multimedia Data within a Mobile Environment. MOMID Deliverable 3, Computer Graphics Center (ZGDV).

[ODP Trading 97] *Information Technology - Open Distributed Processing - Trading Function - Part 1: Specification. Draft Rec. X.950-1 ISO/IEC DIS 13235-1.* (1997). (http://www.dstc.edu.au/AU/research_news/odp/trader/standards.html).

[OMG 95] Object Management Group (OMG). (1995). The Common Object Request Broker: Architecture and Specification; Revision 2.0. OMG TC Document.

[Raposo et al. 97] Raposo, A.B., Neumann, L., Magalhães, L.P., and Ricarte, I.L.M. (1997). Efficient Visualization in a Mobile WWW Environment. Technical Report DCA 001/97, DCA - FEEC - UNICAMP. (ftp://ftp.dca.fee.unicamp.br/pub/docs/techrep/1997/index.html).

[Satyanarayanan 96] Satyanarayanan, M. (1996). Fundamental Challenges in Mobile Computing. *15th ACM Symposium on Principles of Distributed Computing, 1996*, Philadelphia, PA. (http://www.cs.cmu.edu/afs/cs/project/coda/Web/docs-coda.html).

[VRML 97] *The Virtual Reality Modeling Language Specification, ISO/IEC DIS 14772-1.* (1997). (http://vrml.sgi.com/moving-worlds/).

Acknowledgements

This paper is a result of ProBrAl 002/94, a cooperation project between UNICAMP and the TH-Darmstadt, sponsored by CAPES (Brazil) and DAAD (Germany). We appreciate the support granted by these institutions. We would like to thank FAPESP, CNPq, and the Computer Science Dept. (University of Waterloo) for the support of some authors, and also José M. De Martino, Dr. Rüdiger Strack and Jian Zhang for helpful pointers and clarifying discussions.

Intelligent Control of Dynamic Caching Strategies for Web Servers and Clients

Mike Reddy, Graham P. Fletcher
Department of Computer Studies,
University of Glamorgan, Pontypridd,
Mid Glamorgan. CF37 1DL Wales, UK
Tel: +44 (0) 1443 482240
Fax: +44 (0) 1443 482715
Email: mreddy@glam.ac.uk

Abstract: Web pages are cached to reduce network load; various strategies have been adopted which are centred round heirarchies of proxy servers. However, this approach introduces coherence problems. If possible, documents should be kept 'coherent' to prevent delivery of out of date, or 'stale' pages. We suggest that current proxy server and client caching techniques are inadequate for future exponential growth of the Internet as they do not attempt to address the dynamics of document selection and modification. We propose an intelligent dynamic caching technique to model document life histories.This work addresses the coherence problem with particular emphasis on strategies suitable for client browser cache management.

1 Introduction

The phenomenal growth of the World-Wide Web has increased network loads and response times. Various caching techniques have been applied, such as callback, prefetching and validation [Feldmeier 1988]. Callback mechanisms are not appropriate for web objects, which might be cached in many proxies. Prefetching is also unsuitable as current cache hit rates are typically only about 50% [Glassman 1994]. Therefore, pre-emptive document checking may not improve cache performance, as it is difficult to know which objects to prefetch, or when to fetch them. By contrast, validation has been implemented in most proxy servers and browsers. With validation, cached files are time stamped with an expiry date. When further requests for are made, the currency of the cached version is checked. Browsers may be configured to validate the cache every time a document is requested, once per session or never. However, many browsers are notorious for ignoring the spirit of the document expiry, which often results in stale documents being served as current without an adequate warning about their age [Holtman & Kaphan 1995].

When document requests are retrieved, there are three decisions to make for the cache to be maintained effectively:

1. *Which files should be validated?* Traditionally, this has been performed by a combination of using a default expiry time followed by server validation, but this may, in fact, only serve to increase response times and network loads. Any new approach should take into account the 'usefulness' of documents to determine whether they should be replicated, etc.
2. *Which files should be cached?* If there is space and the document is not dynamic (i.e. the result of a CGI request), the file should always cached. If there is insufficient space, there are three simple strategies for deciding which files to cache: 'cache all' removing other files to make space; 'threshold', as before, but

where only files below a certain size are cached; and 'adaptive dynamic threshold' where the maximum file size threshold alters dynamically [Markatos 1996].

3. *Which files should be removed?* The simplest and most common approach to cache management is the LRU (Least Recently Used) algorithm, which removes the most infrequently accessed files until there is sufficient space for the new document. [Abrams et al. 1995]. However, this approach makes no allowance for the 'shelf-life' of cached files, or which documents would be best to remove.

This paper will compare and contrast existing 'semi-intelligent' caching strategies. An intelligent dynamic cache management technique is then proposed, which attempts to improve cache performance by modelling document life histories to determine their usefulness. The use of simulation for the evaluation of such a mechanism is then explored. Finally, the scalability of using this intelligent dynamic cache management technique is examined with regards to both client browsers and proxy servers being implemented as intelligent software agents.

2 Existing Coherence Mechanisms

Several caching mechanisms have been specifically proposed for web documents [Glassman 1994] [Smith 1994]. No existing technique has addressed all three stages of cache management; "Which files to validate, cache and remove?". However, various dynamic, or 'semi-intelligent' techniques have been proposed, which attempt to optimise one or two of these stages. These strategies fall into two main categories: those that address when to validate files, and those that determine which files to cache.

2.1 Expiry-based Cache Management

Heirarchical validation systems are less useful than good time stamping [Bowman et al. 1994]. Better expiry calculation might occur if expiry dates were calculated from the date when the page was last known to be good, rather than from the time that the document was requested. However, this approach is expensive in terms of increased computation and communication overheads, because validity checking and expiry calculation can take several minutes at peak times.

One solution is to have staleness thresholds determined by the manager of a proxy server [Dingle & Partl 1995]. This ensures the speedy propogation of modifications, but is unlikely to be an accurate reflection of when a file is likely to change. Furthermore, this technique is only appropriate for large proxy servers where space is at a premium and documents are often forced from the cache by limited resources. Client caches reflect more accurately the behaviour of individual users, who frequently return to a document within 24 hours or up to a week later, and for whom potential document changes during a particular session will be of no significance.

2.2 Modelling of Document Retrieval

[Pitkow & Recker 1996] have likened caching techniques to models of human memory. Frequency and recency of mental recall have been used to predict the future access to web documents. They conclude that a recency window of one or two days is more useful than frequency for predicting the likelihood of future requests. However, the approach is limited in scope, as it is based primarily upon an empirical model of memory retention [Anderson & Schooler 1991]. Neither are the dynamics of document change included as their study focusses upon large proxy servers which only cache documents for several hours. Another method for determining popular objects has been proposed, which uses the number of requests to gauge a measure of real interest in the document [Dingle & Partl 1996]. However, no evaluation has been performed, so it is difficult to gauge its effectiveness.

2.3 Document Weighting Systems

Coherence has long been a problem in distributed file systems; scoring systems have been used effectively in distributed database systems [Sellis 1988]. However, distributed file system caching may not be appropriate for web page coherence as most accesses are read only. [Bolot & Hoschka 1995] suggest that download time should also be incorporated into a scoring system where lowest weight is used instead of LRU. They propose a weighting metric, which includes the time to last request, document retrieval time, time to live (the header defined expiry date), and document size:

$$w(t_i, s_i, rtt_i, ttl_i) = \frac{w_1 rtt_i + w_2 s_i}{ttl_i} + \frac{w_3 + w_4 s_i}{t_i}$$

where the first term of the RHS is the cost of retrieval against useful lifetime and the second term is the temporal locality. t_i is the time since the last reference, S_i is the size of the document, rtt_i is the retrieval time, and ttl_i is the time to live (the expiry date set by the server or the originator of the document). This compares with the simple LRU approach, where:

$$w(t_i) = \frac{1}{t_i}$$

[Bolot & Hoshka 1995] recognise the difficulty in correctly deriving ttl_i and propose a simplified formula, with suggested typical values of w_1 = 5000 bytes/sec, w_2 = 1000, w_3 = 10,000 byte secs, w_4 = 10 secs:

$$w(t_i, s_i, rtt_i) = w_1 rtt_i + w_2 s_i + \frac{w_3 + w_4 s_i}{t_i}$$

Although the actual performance of the weighting technique was worse than LRU, [Bolot & Hoschka 1995] reported a slight improvement in perceived retrieval time. They measure this using the Weighted Miss Ratio (WMS), which required an additional weighting P, the probability that a file is not in the cache:

$$WMS(t_i, s_i, rtt_i) = P.\left(w_1 rtt_i + w_2 s_i + \frac{w_3 + w_4 s_i}{t_i} \right)$$

where

$$P = \frac{\text{Cache} \quad \text{Misses}}{(\text{Cache} \quad \text{Misses}) + (\text{Cache} \quad \text{Hits})}$$

P is inversely proportional to document size, as small files are accessed more often [Cunha et al. 1995]. Document size and temporal locality are, therefore, important considerations for dynamic caching strategies [Abrams et al. 1995].

Given that rtt_i will be directly proportional to document size, the most significant factor in the formula is S_i. This encourages the indiscriminate caching of large documents in preference to small files. Neither is it likely that rtt_i would be a useful metric without some consideration of the time of transfer or the network performance at the time Furthermore, this technique makes no attempt to model the life expectancy or likely retrieval rate of a web object, ttl_i,. The authors suggest that this is the most significant factor in effective cache management.

3 Intelligent Dynamic Caching

Few of the approaches described above able to modify themselves to changing patterns of network performance. Even when this does occur, the methods often do not use effectively the information available to them on document changes, and the frequency of requests. Rather they make arbitrary decisions as to the life expectancy of documents which are not based upon the actual life history of documents, but merely upon the recency of requests. We propose the use of intelligent caching agents to monitor client behaviour to provide a statistically determined weighting system for cache management based upon document life histories. This approach is achievable with few if any changes in current protocols, as intelligent agents would be expected to work in parallel with existing software.

Two significant factors are the frequency and recency of document requests. The most significant of the two for determining the relative value of documents is the overall frequency of use, which equates to the best use of cache space, irrespective of file size. LRU attempts to predict the likely time between document requests, but only considers recency as its weighting system only uses the most recent transaction. We propose a caching algorithm which effectively does the same, but attempts to predict both the frequency and recency to determine which files should be kept in the cache. An estimate of the mean time to next request (*mtnr*) may be provided by applying exponential smoothing techniques to records of previous requests, as well as the current transaction:

$$mtnr_i = \alpha . t_i + (1 - \alpha). mtnr_{i-1}$$

where t_i is the time since the last reference, *mtnr* is the previous value, and α is the exponential damping factor (typically between 0.1 and 0.3) [Gardner 1985].

The value of α determines the relative importance of the previous frequency and current recency for predicting future behaviour, which will have a direct effect upon performance of the dynamic caching system. Too high a value of α will over-emphasise the recency of documents (the first term of the RHS), which may result in a unrealistic estimate of the likely time between requests. This compares with the underlying strategy of LRU, which is simply a measurement of the most recent transaction, where $\alpha = 1.0$. Too low a value of α, and the algorithm would not react quickly enough to changing circumstances, should the frequency of requests suddenly alter. The weighting metric for determining which files should be removed from the cache is the likely frequency of document requests, which is inversely proportional to *mntr*:

$$w(t_i) = \frac{1}{mtnr_i}$$

4 The WebAgent Simulation

Experimental work is necessary because theoretical models are inadequate for understanding real Internet traffic. We need quantifiable, configurable experiments with large caches, showing realistic network performance and document dynamics. As access to web server log files is often a difficult issue for privacy, security, or logistical reasons, simulation is a necessary tool. This has an added benefit of repeatability, which allows different caching techniques to be measured and compared accurately.

The WebAgent simulator provides an enviroment reproducing file requests (modelling trends in user interest in documents), download delays (seasonal and catastrophic changes in network performance) and document modifications. This enviroment is modelled on the statistics of a web server supplying up to 9000 hits per day. For each caching algorithm the files cached, download times and bytes tranfered are logged. Comparisons between caching systems may be made by simple document hit rate (DHR), byte hit rate (BHR), which are simple measures of how many files (or bytes) are retrieved from the cache. A more accurate metric is the perceived retrieval rate, *PRR*, defined as the total number of bytes delivered divided by the time spent retrieving files which were not cached:

$$PRR = \frac{\sum_{i=1}^{n} S_i}{\sum_{j=1}^{m} t_j}$$

where n is the total number of files delivered, S_i is the size of file i, m is the number of files which were not cached and t_j is the time to download a file j.

4.1 Simulation Results

[Fig. 1] shows the perceived retreval rate for LRU and the agent based cache system. From these results it can be seen that the use of a damped adaptive measurement for the mean time to next request consistantly outperforms the LRU algorithm. The agent reacts quickly to changing network performance, while still maintaining a greater perceived retrieval rate for document delivery.

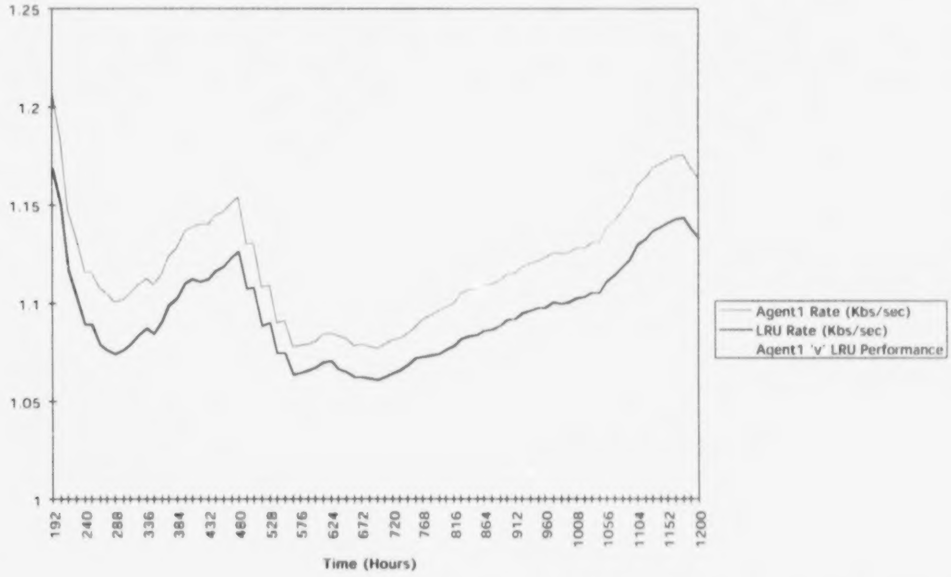


Figure 1: Comparison of Agent and LRU perceived retrieval rates

5 Conclusions and Future Work

In this paper, various techniques for management of client and server based web caches have been examined. The case has been made for an intelligent agent, which models the usefulness of web objects by evaluation of document life histories. We propose such a system, which though less than optimal, still shows an improvement in the handling of web objects over existing techniques, such as LRU. The WebAgent simulation has reproduced results which suggest that the frequency of requests for a documen, rather than file size, is more relevant to the management of web caches. Agent defined estimates of document request rates, can significantly improve performance. Furthermore, the dynamic nature of our approach, should provide ever improving performance, and a system which can resolve itself to frequently variable network use and performance. Although these techniques have been aimed primarily at client caches, the authors believe that they are appropriate and scalable to proxy servers.

A number of future improvements have been identified for the WebAgent simulation, which will serve to both improve the accuracy of the simulation. The following future work will be based upon the simulation:

1. Investigate the use of agent modelling to predict modification times, to allow prediction of likely stale documents
2. Extend the simulation to include multiple, geographically distributed servers, with more realistic weightings for network performance.
3. Implement pre-fetching algorithms which can make use of 'off peak' times to maintain cache coherency for frequently used documents.
4. Investigate the use of pre-emptive distribution of documents to mirror servers, and maintenance of distributed document modifications.
5. Develop server-based agents to analyse geographical trends in user access to popular documents, to improve performance on distributed mirror servers.
6. Construct an intelligent dynamic caching agent for an existing browser, to evaluate performance of an actual user on a real network.

6 References

[Abrams et al. 1995] Abrams, M., et al. (1995). Caching proxies: limitations and potentials. *Proceedings of the 4th World Wide Web Conference*. Boston, MA. 119-133.

[Anderson & Schooler 1991] Anderson, J., & Schooler, L. (1991). Reflections of the environment in memory *Psychological Science*. 2 (6), 192-210.

[Bolot & Hoschka 1996] Bolot, J-C., & Hoschka, P. (1996). Performance engineering of the World Wide Web: Application to dimensioning and cache design. *Proceedings of the 5th International World Wide Web Conference.* Paris, France.

[Bowman et al. 1994] Bowman, C.M., et al. (1994). The HARVEST information discovery and access system. *Proceedings of the 2nd World Wide Web Conference*. 763-77

[Cunha et al. 1995] Cunha, C., et al. (1995). *TR-95-010: Characteristics of WWW client based traces.* Boston, MA. Boston University.

[Dingle & Partl 1996] Dingle, A., & Partl, T. (1996). Web cache coherence. *Proceedings of the 5th International World Wide Web Conference*. Paris, France.

[Feldmeier 1988] Feldmeier, D. (1988). Improving gateway performance with a routing table cache. *Proceedings of INFOCOM'88'*. 298-307.

[Gardner 1985] Gardner, E.S. (1985). Exponential smoothing: the state of the art. *Journal of Forecasting*. 4. 1-28

[Glassman 1994] Glassman, S. (1994). A caching relay for the World Wide Web. *Computer Networks and ISDN Systems*. 27 (2), 165-173.

[Holtman & Kaphan 1995] Holtman, K., & Kaphan, S. (1995). *Problems with the expires header*. <http://www.amazon.com/expires-report.html>

[Markatos 1996] Markatos, E.P. (1996). Main memory caching of web documents. *Proceedings of the 5th International World Wide Web Conference*'. Paris, France.

[Pitkow & Recker 1996] Pitkow, J.E., & Recker, M.M. (1996). *A simple yet robust caching algorithm-based on dynamic access patterns*. <http://www.vuw.ac.nz/~mimi/www/www-caching/caching.html>

[Sellis 1988] Sellis, T. (1988). Intelligent caching and indexing techniques for relational database system. *Information Systems*. 13 (2), 175-185.

[Smith 1994] Smith, N.(1994). What can archives offer the world wide web. *Proceedings of the First International World Wide Web Conference*. Amsterdam, Holland.

A Database Architecture for Web-Based Distance Education

Daniel R. Rehak
Professor, Civil and Environmental Engineering
Chair, University Multimedia Courseware Committee
Carnegie Mellon University, Pittsburgh PA 15213, USA
Email: rehak@cmu.edu

Abstract

The goal of the Carnegie Mellon Online project is to build an infrastructure for delivery of courses via the World Wide Web. The project aims to deliver educational content and to assess student competency in support of courses across the Carnegie Mellon curriculum and beyond, thereby providing an asynchronous, student-centered approach to education. The system centers on a formal data model, supported by commercial database technology. This presentation provides a technical overview of the system and its features and discusses our initial use of it in our large introductory courses.

1. Introduction

Developing Carnegie Mellon Online was initiated to move our introductory computing competency course to the Web and offer it as a distance education opportunity to our incoming students. As a result, a general system for Web-based course delivery was developed and has been in use since Summer 1996.

We have built a course-independent, database-driven, student-centered, Web-based delivery mechanism for education and training. The system has been used both on campus and to support distance education, and is designed to be scalable to both large classes and courses with different structures and needs. The system aims to deliver educational content and to assess student competency in support of courses across the Carnegie Mellon curriculum and beyond, thereby providing an asynchronous, student-centered approach to education.

Based on the initial success with the project, and the opportunities to bring TEL (Technology Enhanced Learning) aids to more courses, work is continuing on the system, and additional courses are being developed and deployed.

2. Technical Description

Carnegie Mellon Online is a system for course-independent, Web-based delivery of educational, training and assessment materials. The system generates customized content (e.g., assessments, feedback) for each individual student and tracks the student through a course while enforcing course-specific rules and policies. The system model clearly separates course educational content, course structure, course policies and individual student records.

The system is designed to be scalable to handle large numbers of courses and students, and to provide sharing of educational content across multiple courses. It includes appropriate network security and authentication. On the client side, the student needs only a Web browser and an Internet Service Provider.

All course descriptive elements and student records are stored in a database. Individualized content is dynamically computed and delivered to the student via the Web. Courses are represented in a formal model as a structured collection of the elements of a course (e.g., instructional modules, exams, tutorials, assignments), where these elements of course content are shared across courses. The model of content is augmented with information on how courses are taught and their operational rules and policies (e.g., prerequisites, grading criteria). The entire system is data driven; only basic course-independent modeling concepts are represented directly in the system model. All content and course operations are declarative information in the data model, processed by the generic system engine. The overall system structure is shown in Figure 1.

2.1 System Architecture

The system is structured into six major elements:

- **Course delivery:** providing educational content and course status information to students via the Web and tracking their progress through a course via the database.
- **Course management and administration:** managing students, classes, staff, sections, etc., independent of actual courses and course material (e.g., moving a student from section x to section y). Course management functions are accessed via the Web.

- **Student administration:** grading, providing information about individual students to instructors, etc. Again, via a Web interface.
- **Content authoring:** preparing models of a course, its structure and its policies, and entering them into the database. Also, via a Web interface.
- **Operations:** daily operations and maintenance procedures (e.g., backups, system monitors, course announcements).
- **Content:** storage for educational content and operational rules for individual courses or elements of courses delivered via the course delivery system.

Wherever possible, course-specific features are modeled declaratively in the database, either as elements of the course and system model or as elements of course content. Most of the code contains no course-specific features.

2.2 Course and Content Model

A key feature is the model of courses within the system. Courses are modeled as a multi-level collection of different types of course elements. The general approach is to provide an arbitrary number of levels of information. Elements can be shared at any level within one or more courses (see Figure 2).

At the top of the hierarchy are a collection of **components** and **subcomponents**. These are typically the units which are major portions of a course. There are an arbitrary number of such units used to describe any course, with an arbitrary number of elements within each layer. These elements essentially describe how more specific content elements are combined into a course. Associated with each of the top-level units is the information describing policies and course operational data such as grading criteria, access and authentication rules, prerequisites, etc.

The units at the lowest level of the component hierarchy are denoted **modules**. Below these component elements, each of the units of a course consists of a set of **content** elements (denoted **devices**). Content elements are classified by their role. Element roles include **learning**, **assessment** and **feedback**, along with special system operational and administrative content elements. Each module unit may have an arbitrary number of each different type of these content device elements. These content elements are composed of a set of **parts**, where each part represents a single point of interaction between the student and the course delivery system.

Devices are the primarily educational content elements of a course. Each type of device has an associated set of operational rules. For example, learning devices are information delivered to the student (e.g., lectures, syllabus) without need for student input; they may be individualized but generally are shared content for all students. Assessment devices involve presenting assessment information (typically individualized), getting results from the student, grading, and using the grade results to control the student's next step in the course (e.g., passing an exam provides a prerequisite for the next topic). There are specific models of the information associated with each type of device (Figure 2 shows more details of the model for assessment devices).

Each course is a collection of the different elements of course content and descriptions from the database. The course author selects and combines them into a single course. A set of staff, teaching sections, schedule and other administrative information is associated with the content, and this is combined with a list of registered students to fully instantiate a course for Web delivery.

While the information in the model appears to be hierarchical, it is not. Each level is described separately and named globally, and elements in any level are unordered collections of elements from the lower levels. Various parameters and course rules can be entered at many of the levels and the system can use the data from one level to override the values for the same attribute from a lower level when the individual elements are combined into a specific course.

In addition to the modeling components which describe a course, the elements of the system itself are represented in the same structure, e.g., there are **system** elements at each level in the model. Similarly, there are elements for operations and management at each level in the model. The actual data modeling is generalized to accommodate such elements, i.e., at the kernel level, the system itself represents courses and internal operations with the same representations and data structures.

2.3 Pedagogical Elements

A strength of the current implementation is its ability to handle assessments. The course and content models contain a number of specific features to represent and process assessments (e.g., examinations). For example, the system can generate custom examinations for individual students based on a number of different criteria and parameters, including their prior work and results. Examinations are also defined in terms of the topics and types of questions (multiple choice, matching, value, task instructions, etc.), and individual questions include details of grading policies, instructions to the grader (when graded by an instructor), and feedback for the student. Generated examinations are delivered to the student, and the student's work is submitted and either graded automatically or sent to an instructor for personalized grading. In either case, detailed, customized feedback is generated and returned to the student. The amount and type of feedback are determined by the policy rules associated with the assessment device and the course.

By tracking student progress versus course requirements, the system can present a customized view of a course to the student, e.g., only work for which prerequisites have been completed is presented. The interface design is task-focused and student-centered, presenting only appropriate information to the student at each step. This eliminates the need for the student to search through the course and Web to find appropriate elements to make progress towards completing the course.

Courses built to date incorporate learning and teaching elements which are hypertext and are rich in assessment materials. This is not a limitation in the system design, but rather in available content. Multimedia, simulations and active tutors can be added and are under development. These enhancements can be accommodated by the existing course and content model. Each of these items can use the state information about student progress and the overall course model to provide customized views and student-specific content within the overall course structure and operational rules.

2.4 Database

The entire information model is stored in a database. Modeling and representation of course-specific information is clearly separated from course-independent information.

The major collections of information within the database are:

- **People**: information on students and teaching staff, independent of their association with a specific course.
- **Courses**: registration-type course information, such as schedule, and associated staff and students.
- **Content:** course content such as questions, answers, learning materials, tutorials, and course evaluations. Content is not associated with a specific course but can be shared across courses. Similarly, it is not associated with an individual student, but is instantiated as necessary for the students in a course.
- **Course Models:** descriptions of selected elements of content which comprise a course, along with related course policies. This includes modeling a course in terms of its components and modules, and associating specific content with the elements within the modules. The course models also include the internal system models in the same framework.
- **Course Instances:** selection of a course model and association with people (students, staff) and ccurse (registration) information, customization of the content (e.g., exam generation) for each student.
- **Student Records**: tracking of the progress of individual students through courses (including a full record of student access to their personal content). The record information also includes system records in the same format.
- **Media:** media files, HTML, graphics, programs, etc., used by either the system or the courses.
- **HTML:** dynamically generated HTML saved for later delivery; stored in the database for indexing and security.
- **Code:** database procedural code.

2.5 Interface

The Web interface is specific to each individual course, in terms of its organization and form; designers are free to pick a look and feel which is appropriate for specific courses. There is a single link between the interface and the course delivery system, and the link structure is course independent. Thus a different course interface could be installed and the rest of the system would remain unchanged. A new look is developed by providing new image files and screen layout. How the interface is related to the delivery of course content and system operation is also encoded in the description of the course model.

While the system is designed to be student centered, we have been constrained by the commonly deployed Web technologies available to our distance students (courses are offered to our students around the world) and thus have limited the amount of state information portrayed via the interface. The types of interfaces which we desire to deploy require both Java and Javascript, and we have not found these technologies to be robust enough yet to deploy and support across all of our target platforms. A more advanced Java/Javascript interface is under development. This interface will maintain more information on the student or client side of the Web connection, and will not only provide active feedback, but will also focus the student on only the appropriate actions at any point in a course. Thus the system will be better able to dynamically lead a student through a course.

2.6 Technology Base

Carnegie Mellon Online is a custom system constructed from standard components. The components and system conform to Internet and other standards whenever possible:

All information is stored in a commercial relational database (Oracle V7). The course and content model is represented directly within the database. A commercial Web server (Oracle WebServer 2) provides the Web interface to the system and to the database. Most of the system consists of code (Oracle PL/SQL) to provide database access via the Web, to compute custom content, and to maintain the model of courses and students. Security is provided with SSL Web transactions and site authentication is implemented with Kerberos login. Client state is maintained with secure cookies.

Students need only a basic Web browser which supports frames, e.g., Netscape 2 or above. Interfaces must work at 640x480 with 256 colors, and response should be reasonable over a 14.4kbaud dialup connection. Courses are delivered on Macintosh, Windows and Unix clients.

Our production system is deployed on a Sun server. Hardware resources dictate the size of a course which can be represented in the database and the system response rate. The current production server is a dual processor, with dual I/O subsystems, 256MB of primary memory and 20GB of disk. The configuration is designed to provide subsecond Web/database response under normal peak loads of 50 users all requesting information simultaneously. We have development and test systems which operate on NT PCs, and a small course can be delivered using a small desktop or notebook PC.

3. Use

Carnegie Mellon Online was initially developed to support our introductory computing competency course (CSW). Starting with a pilot in the Summer of 1996, this course is now fully supported and Web delivered. Many of the features of the system architecture were driven by the demands of this course. Additional courses and pilots have been or are under development.

3.1 Computing Skills Workshop

Computing Skills Workshop (CSW) is an entry-level, one credit course, required of all of our students, covering how to effectively use computing throughout the curriculum. Its aims are to: (1) introduce new students to Carnegie Mellon's computing facilities; (2) ensure that all students have a baseline of conceptual knowledge and practical skills using a variety of productivity software (word processing, spreadsheets, etc.); and (3) equip students to make effective use of computing in the service of their other courses.

CSW traditionally was a large lecture and laboratory course operating in an assembly line process of feeding students through a formal schedule of topics, assignments and an end-of-semester examination. The course is taught both semesters, but most incoming students are scheduled for the Fall semester, with a typical Fall enrollment of 1300+ students. Labs and lecture sessions are led by undergraduates (a staff of about 50), and the course runs in two dedicated computer clusters (25 seats each) from 9:00AM to 8:00PM every day.

A plan was developed to make the course modular and self-paced, letting students take any of the modules in any order, at any time, and to support it with Web-based delivery of materials, submission of work and electronic grading. By utilizing the Web, the course would be freed from the time and place constraints of our campus and schedule and many of the course management functions could be automated. While the use of the Web is essential, the goals are driven by the desire to reshape the model of the course, not how technology supports it. Benefits include enabling students to complete the course early, providing more feedback and freeing staff resources to spend more time with students who need individual attention.

First-order benefits include: (1) students can concentrate on CSW in the summer between high school and college when there are fewer competing demands and distractions; (2) with CSW successfully behind them, students will have more time for academic demands and social opportunities once they arrive on campus; (3) students will arrive on campus with certified computing skills, prepared to apply those skills to academic courses. Second-order benefits of creating a general system include: (1) providing a powerful model of Web-based education which we can use to build additional materials for full courses or modules; and (2) creating a framework in which to imbed ongoing work in cognitive tutoring, just-in-time education, automated curriculum design, and educational evaluation.

3.2 CSW Online Structure

CSW Online is primarily a system for assessment, not teaching. The content model is rich with assessments but thin on learning materials. Web materials are complemented with traditional texts. On campus, students can attend lectures to learn about the different topics in the course.

The course itself currently covers several topics or modules: word processing, spreadsheets, email, basic Unix and Emacs, and networking (Internet, File Servers, Libraries). Once a student thinks she is ready, she can take a competency examination for a topic. Passing the course requires all modules be completed.

CSW Online presents the students with a choice of activities for each module. First they can read about the competency topics included in the examination on the topic. This is the information they need to know to successfully pass the examinations. There are limited on-line learning materials, but the topic information is indexed to the textbook (if copyright for the textbook materials were available, the information would be linked to on-line books).

For each topic, there are practice and graded examinations. In the current course (Fall 1997), the examinations are situated tasks for which the student is asked to take a set of resource files and a set of instructions and produce a new document. We have explored different options for the examinations, including multipart examinations that included automatically generated and graded sets of questions to test declarative knowledge.

We have also explored different policy options, including a fully self-paced course versus a directed schedule with due dates, and have used different mixes of take-home versus in-class graded worked. All of these variants have been built by simply providing different course models on top of a common set of content materials. The current course is actually four different combinations of course models. One option is the choice of PC- versus Macintosh-specific educational content. The other is self-paced versus directed course policy. All are delivered from the same database, and with the same interface. The content is individualized for students based on their course registration.

3.3 CSW Deployment

CSW was initially developed as a pilot over the summer of 1996. We built the system and delivered two modules (40%) of the course to incoming students working at distance. From an initial pilot group of 250 students, we had approximately 100 complete a portion of the course during our month long pilot.

The summer pilot was a technical success, and in Fall 1996, we delivered two modules of CSW via the Web to a class of 1400 and traditional delivery (e-mail, bboard, etc.) was used for the rest of the course. For Spring 97 we completed the creation of all content and had all modules available for Web delivery and supported 300 students.

CSW Online is now an option for all incoming students in the summer before their enrollment. In Summer 1997, the course was available as a distance education option, and over 250 incoming students worked on the course from around the world, and some had completed it within days. In addition to the incoming students, the course was offered on campus to upper classmen who had not finished it, to students in our summer precollege program and as staff training. Total enrollment was 400. In Fall 1997 we will have 1300+ students in the course, but those who worked on it during the summer will need to complete only one new email module which requires software only available on campus.

3.4 Other Courses and Plans

Over the Summer of 1997 we offered a small placement examination for our incoming calculus students. Traditionally, all of our calculus students are given a paper examination, delivered by mail. The results of the examination and a survey are used to place the students into the appropriate course in our calculus sequence.

In the pilot, the examination and survey were converted into a small Web-based course. The course included a background survey, a practice examination used to familiarize students with Web-based assessments, and the complete placement test. The course was designed to be completed in a single Web session, and the students were given immediate feedback on their performance and placement.

Our Computer Science Introductory Programming course (C++) is being developed for on-line delivery. This is a large, intense, mainstream academic course with 3 units of credit. Passing the course with a grade of B or above is a prerequisite for our upper-level CS courses.

A pilot course was offered at a distance to selected students in the incoming class during the Summer of 1997. This course includes online lecture materials, problem sets, practice programs and examinations for each course topic. It includes a large body of content, over 1000 questions for examinations and over 300 programming exercises. The complete pool of materials is significantly larger than would be presented to any one student. We are also investigating Web-based compilation to support students who do not have the necessary compiler on their computers, and automated grading of programming assignments. Building the course involves defining the course model and loading it, along with the course content, into the database. A new interface structure is being designed, but there no system level code or data modeling changes are required.

Completing Introductory Programming requires passing a single mastery exam. Currently, about 1% of the 1000 students enrolled each year can successfully complete the examination without taking the course. By offering the course at a distance to our incoming students, we believe that 10% to 30% of them will be able to complete the mastery exam within the first two weeks of the semester. The students will benefit from completing the course early, permitting them either to take another course in its place, or to devote more time to other activities. We do not plan to reduce the staff, but rather redeploy the resources to provide more one-on-one support.

Other courses are being developed for deployment during AY 97-98, including one on Engineering Economics and one on Art History. In addition, the project team has received inquiries from faculty wanting to develop courses in Introductory Biology, Introductory Chemistry, and Statistical Reasoning. We are also continuing to add features and capabilities to the system to better support courses and their operations.

Acknowledgments

Elements of this report were extracted from prior documents describing CSW, the CSW Online Proposal, and the 96 Summer Pilot report. In particular, credit for some of the words above goes to Chris Thyberg, Director of CSW; Phil Miller, Director of Introductory Programming; and Diana Bajzek, Director of TELab.

Support for Carnegie Mellon Online has been provided from the Office of the University President. The project is under the operational control of the Division of Computing Services.

More Information

For more information, visit our Web site: http://online.web.cmu.edu/ or contact us via e-mail at: online+@cmu.edu

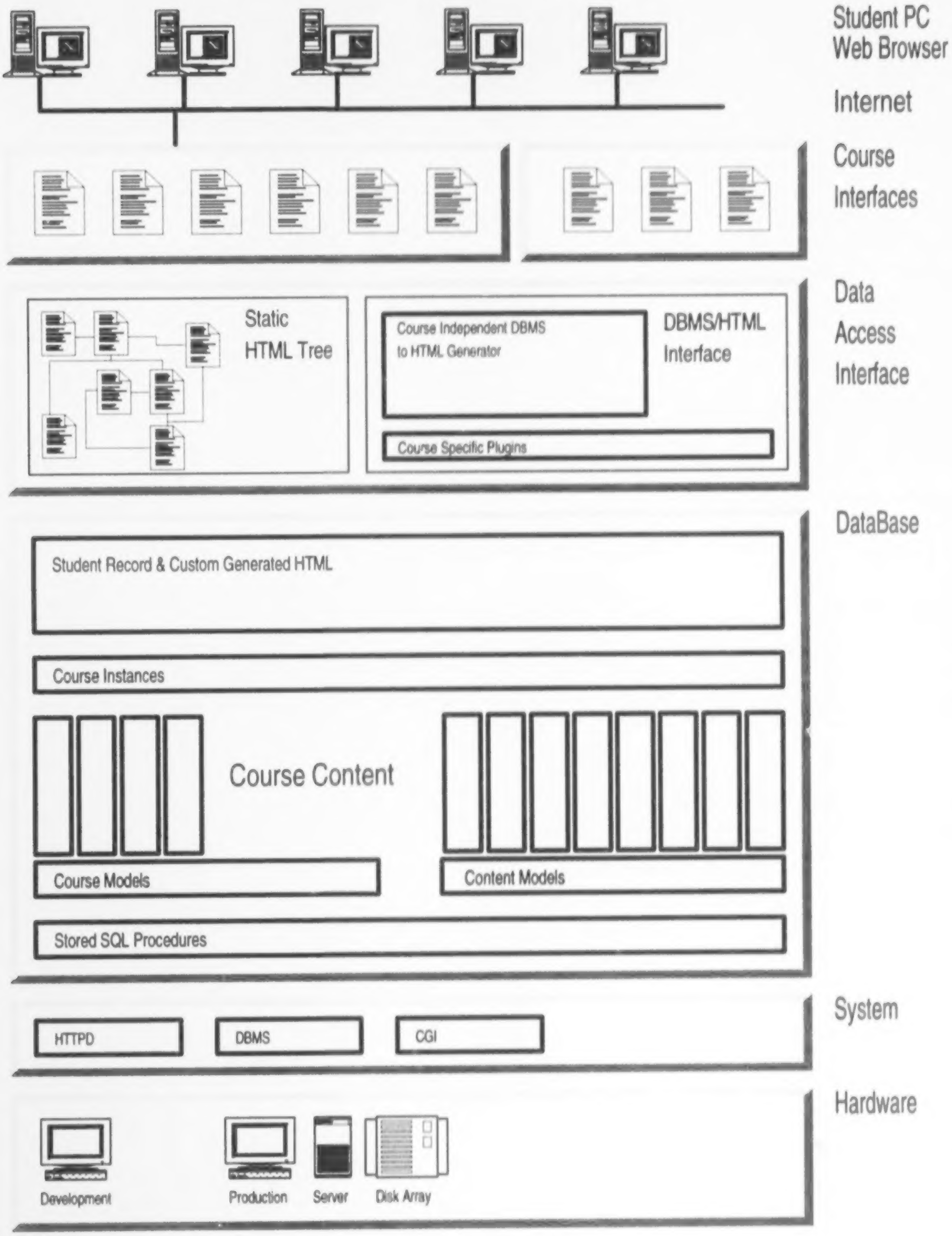

Student PC
Web Browser
Internet
Course
Interfaces
Static
HTML Tree
Course Independent DBMS
to HTML Generator
DBMS/HTML
Interface
Course Specific Plugins
Data
Access
Interface
DataBase
Student Record & Custom Generated HTML
Course Instances
Course Content
Course Models
Content Models
Stored SQL Procedures
System
HTTPD
DBMS
CGI
Hardware
Development
Production
Server
Disk Array

Figure 1: Carnegie Mellon Online System Structure (patent pending)

BEST COPY AVAILABLE

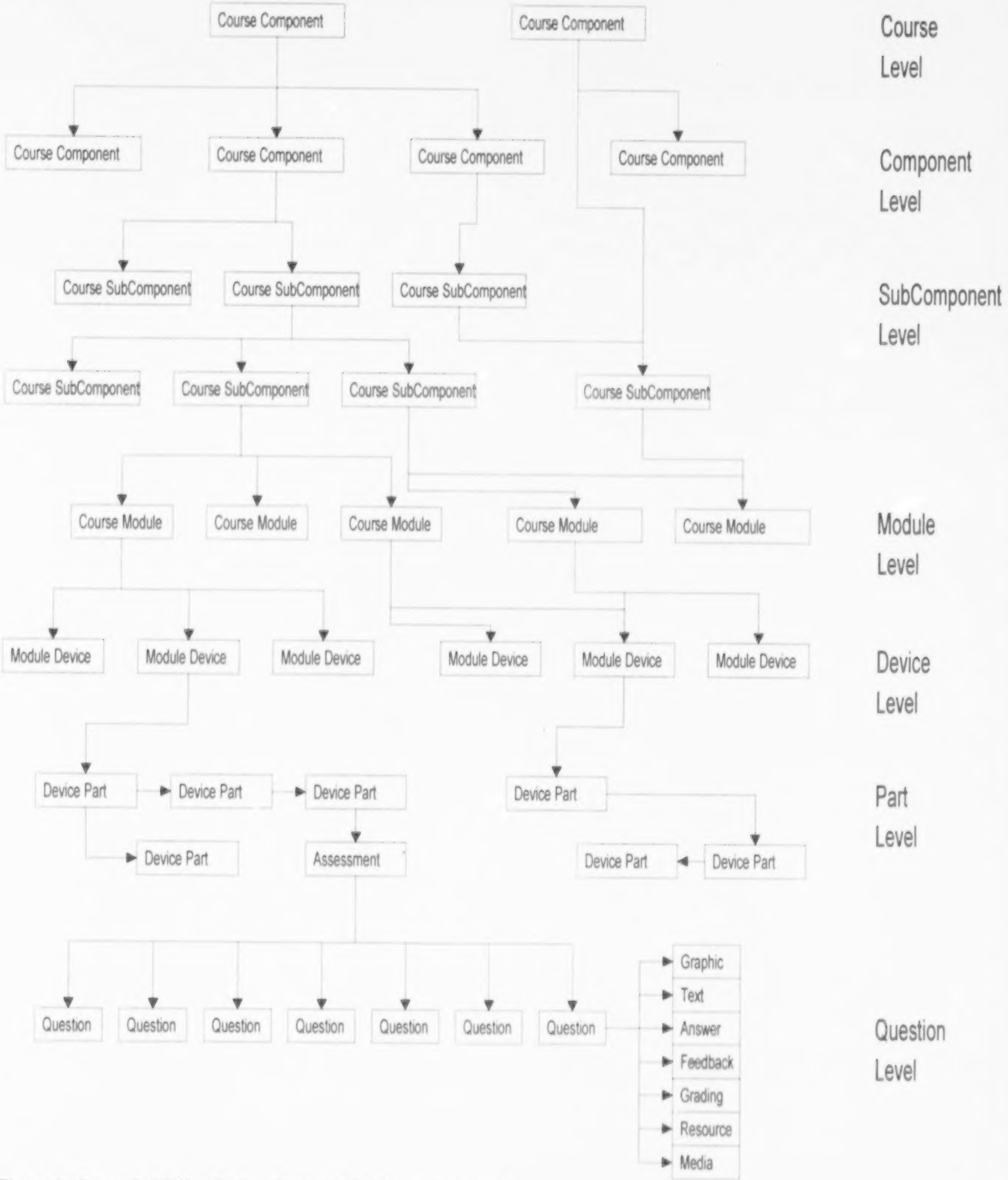

Course Component
Course Component
Course Level
Course Component
Course Component
Course Component
Course Component
Component Level
Course SubComponent
Course SubComponent
Course SubComponent
SubComponent Level
Course SubComponent
Course SubComponent
Course SubComponent
Course SubComponent
Course Module
Course Module
Course Module
Course Module
Course Module
Module Level
Module Device
Module Device
Module Device
Module Device
Module Device
Module Device
Device Level
Device Part
Device Part
Device Part
Device Part
Device Part
Device Part
Assessment
Device Part
Device Part
Part Level
Graphic
Text
Answer
Feedback
Grading
Resource
Media
Question
Question
Question
Question
Question
Question
Question
Question Level

Figure 2: Carnegie Mellon Online Course Model (patent pending)

BEST COPY AVAILABLE

Paper Interfaces to the World-Wide Web

Peter Robinson, Dan Sheppard & Richard Watts
University of Cambridge Computer Laboratory
Cambridge CB2 3QG England
{peter.robinson,dan.sheppard,richard.watts}@cl.cam.ac.uk

Robert Harding & Steve Lay
University of Cambridge Department of Applied Mathematics and Theoretical Physics
Cambridge CB3 9EW England
{r.d.harding,s.w.lay}@damtp.cam.ac.uk

Abstract: This paper reports on ways of using digitised video from television cameras in user interfaces for computer systems. The *DigitalDesk* is built around an ordinary physical desk and can be used as such, but it has extra capabilities. A video camera mounted above the desk, pointing down at the work surface, is used to detect where the user is pointing and to read documents that are placed on the desk. A computer-driven projector is also mounted above the desk, allowing the system to project electronic objects onto the work surface and onto real paper documents.

This paper describes a particular application in which the system is used to provide access to the World-Wide Web. A WWW page can be printed on paper and then placed on the digital desk and animated: when a link is selected with a pen, the corresponding link in the original HTML document is followed and the resulting page projected onto the desk.

Background

Recent developments in electronic publishing have shown the value of hypertext both for documents on CD-ROM and also for on-line presentation through the World-Wide Web. Computers endow electronic documents with powerful new facilities, leading some to believe that electronic media will soon replace conventional media completely. The trouble is that people like paper. It's portable, tactile and easier to read than a screen; in fact, computers now generate far more paper than they replace.

At the same time, developments in computer hardware have greatly reduced the cost of attaching television cameras to computers. They have moved from being an expensive peripheral for specialists to a price comparable with a monitor; further developments in technology will soon make the cost similar to that of a mouse. This raises the question of what new techniques will be appropriate when every computer routinely includes video input, possibly from several cameras.

Over the past few years, the University of Cambridge Computer Laboratory and the Rank Xerox Research Centre in Cambridge (formerly EuroPARC) have collaborated on research into the use of video in user interfaces [Robinson 1995, Stafford-Fraser 1996, Stafford-Fraser & Robinson 1996, Wellner 1993, Wellner 1994]. Computers 'watch' users at work and infer commands from gestures involving pens and paper. This is not *virtual reality* where the user is immersed in a totally synthetic, computer-generated environment, often donning a special headset and even clothes; this is *augmented reality* where the computers operate through everyday objects in the real world, enhancing them with computational properties.

Such a system requires the computer to monitor activities and to deliver its contribution as conventionally as possible, suggesting the use of video and, to a lesser extent, sound for input and output. Of course, this merely reflects normal practice. We are used to pointing to interesting parts of documents and commenting on them; electronic enhancements should operate in the same way.

At the same time, electronic, multi-media publishing has emerged as an alternative to conventional publishing on paper. The World-Wide Web and versions of reference books and fiction published on CD-ROM can enhance their conventional counterparts in a number of ways:

- They offer elaborate indexing, glossaries and cross-referencing.
- They allow non-linear progression through the text.
- Sound and moving images can be added.
- Sections can be copied into new documents.

However, screen-based documents have a number of disadvantages:

- People find screens harder to read than paper.
- Electronic bookmarks are less convenient than bits of paper.
- Adding personal notes to electronic documents is difficult.
- Writing, editing and proof-reading a non-linear, multi-media document is still a specialised and difficult task.

We have been investigating ways of resolving these difficulties by publishing material as ordinary, printed documents that can be read in the normal way, enjoying the usual benefits of readability, accessibility and portability. However, when observed by a camera connected to a computer, they acquire the properties of electronic documents, blurring the distinction between the two modes of operation and giving a richer presentation that that afforded by either separate medium.

Our initial experiments have applied this technology to computer-assisted learning [Harding et al. 1997]. Earlier work with Computer Illustrated Texts [Harding & Quinney 1990] supplemented printed books with software that was an integral part of the educational package but which had to be run separately. The two parts can now be united and a number of applications have been investigated. Separate papers discuss the presentation of mixed-media documents [Robinson et al. 1997a] and the internal architecture of our system [Robinson et al. 1997b].

This paper describes a particular application in which the system is used to provide access to the World-Wide Web [Berners-Lee et al. 1994]. A web page can be printed on paper and then placed on the digital desk and animated: when a link is selected with a pen, the corresponding link in the original HTML document is followed and the resulting page projected onto the desk. Sections of the documents can be captured in electronic form, edited, printed and animated in the same way.

Architecture

The overall architecture of the animated paper document system is shown below [Fig. 1]. The system is written in Modula-3 [Nelson 1991], a high-level systems programming language whose object model has been extended to operate in a distributed environment [Birrell et al. 1993]. The principal components are as follows:

The Registry

At the core of the system is a *Registry* which maintains the association between electronic documents and their printed variants. It stores the image of each active document and the code of any interactions required for the document, together with cross references between these and further indexes to identify them. In the context of WWW documents, these correspond to links to other URLs, but the facilities allow much more general forms of interaction.

In the current implementation, the code implementing the interactions has to be linked in to the system. However, this is just a temporary measure. A better long term solution would be to store complete programs as Java applets [Arnold & Gosling 1996] or Obliq oblets [Brown & Najork 1996] which are more amenable to dynamic loading for remote execution. This would also simplify the handling of Java embedded in documents handled by the system.

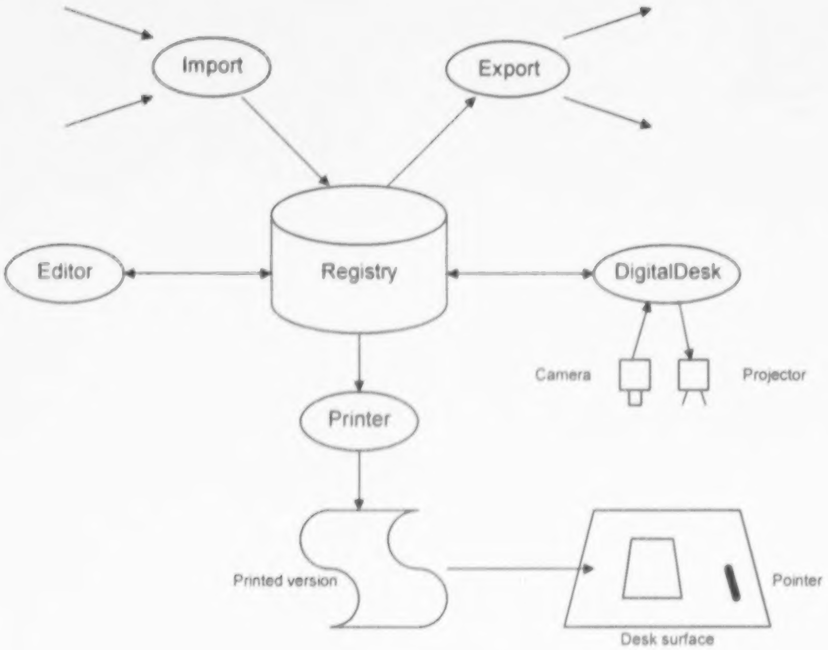


Figure 1: Animated paper document framework.

The registry is accessed via a set of adaptors that allow the database to be built and edited, imported and exported to other forms of hypertext, and for documents to be printed and animated on a DigitalDesk. The following are relevant for processing WWW documents.

Import

Conventional hypertext can be absorbed into the animated paper document system; paper access to the World-Wide Web is possible through such an adaptor. Given a uniform resource locator (URL), the adaptor captures the information currently on the associated Web page in the registry. This includes the URLs of any links embedded within the page.

An HTML parser breaks the document into blocks of text (usually paragraphs, but at a finer grain where there are links) and images. These are then rendered as PostScript and the positions and content of the links recorded. All this information is kept in the registry. The page can then be printed simply from the PostScript, with further embellishment to assist subsequent page recognition.

Editing

Documents in the registry can be edited with a fairly conventional WYSIWYG editor. Text and diagrams are entered and amended in the usual way. However, it is also possible to mark areas of the document as hyperlinks and associating interactors with them. These are recorded as references to the associated code.

One version of the editor actually operates on the DigitalDesk, which means that text, diagrams and interactors from other printed documents can be copied into the new document. If the other printed documents are active documents known to the system, this copying is entirely digital, just as it would be in a conventional word processor. However, text and pictures can also be copied from conventional printed documents by using the overhead camera to capture an image and passing any text through an optical character recognition system.

Printing

Another adaptor prints out documents from the registry onto paper so that they can be used for direct interaction on the DigitalDesk. The printed documents are annotated with marks in their corners to facilitate recognition and location on the desk top, and are also have a unique identifier printed in an OCR fount.

Once the document has been printed, its contents are retained in the registry as an immutable copy of its structure for future interaction. This allows the paper to continue working in the same way even if its electronic original is edited. However, any URLs referred to in the electronic version will have been remembered and will be followed when the paper version is animated. The contents of the pages identified by such URLs can change in the usual way.

DigitalDesk

The DigitalDesk actually animates the paper documents. This involves recognising that a page printed by the system has appeared on the desk, determining its position, reading its unique identifier and locating any interactors. A transformation is then set up between the page representation stored in the registry and physical co-ordinates on the desk top. The printed document thus becomes part of the projected window system. In particular, any active links are highlighted by projecting a red background over them. For a document originating on the Web, these correspond to links in the original HTML.

A pen with a light-emitting diode in its tip is used for pointing. This is recognised by the camera system and converted to co-ordinates using a transformation calculated by occasional registration. It would be possible to use a conventional graphics tablet, but the light pen has the advantage that it works perfectly well over a stack of paper on the desk. The events are passed back through the window system and interpreted using information in the registry. For a URL, this involves opening a new projected window on the desktop and displaying the contents of the associated page in it. The Modula-3 window system, Trestle [Manasse & Nelson 1991], and its user-interface toolkit, FormsVBT [Brown & Meehan 1993], include a window primitive that acts as a WWW browser, so this is straightforward.

Export

The interactions afforded by animated paper are considerably richer than straightforward HTML but if a document is sufficiently simple, it can also be exported as HTML.

This involves scanning the image of the page from top to bottom, left to right and emitting text or images as appropriate. When a page is published in this way, a series of HTTP PUT commands are sent to the WWW server which is going to hold the page. One of these is for the HTML of the page itself, the others are for the richer features of DigitalDesk documents that can not be translated into conventional HTML. These can be recovered either through another DigitalDesk or by a suitably extended WWW browser.

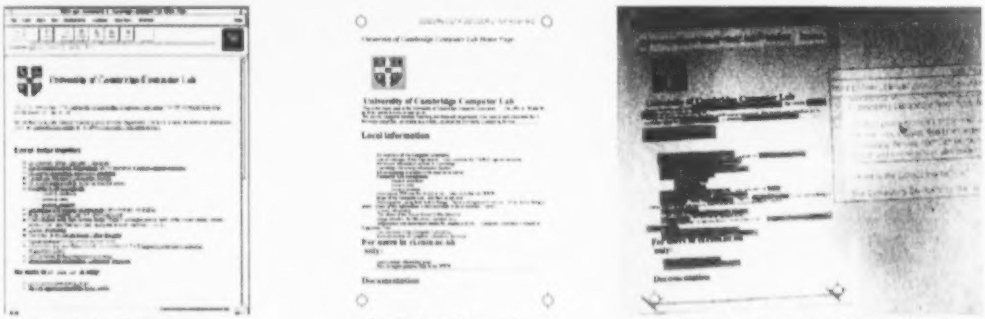
Operation

The pictures below [Fig. 2] show the system in use as a World-Wide Web page passes through the stages just described:

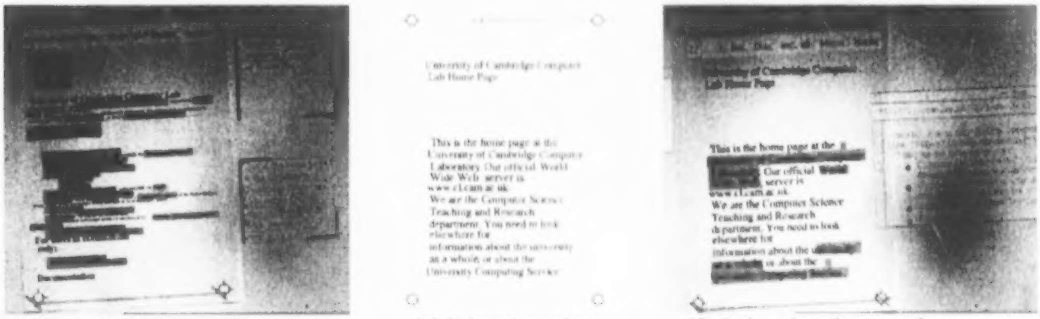
(a) The Computer Laboratory's WWW home page is displayed by a conventional browser.

(b) This is imported into the animated paper document system's registry and printed on paper with extra decorations to assist recognition.

(c) When this is placed on the DigitalDesk it is recognised and active areas of the document are illuminated by projected highlights. One of the links has been followed and the contents of the associated URL are being projected onto the desk through a browser running in a separate window.



(a) Original Web page. (b) Printed version. (c) Animated on the DigitalDesk.



(d) Deriving a new document. (e) Printed version. (f) Animating the new document.

Figure 2: Paper access to the World-Wide Web.

(d) The editor is invoked and sections are copied from the paper document into a new electronic document projected onto the desk. This uses conventional copy-and-paste but works from a paper document to an electronic one. Existing links can be copied and new links added.

(e) The new document is printed with the standard decorations.

(f) The new paper document can also be animated on the DigitalDesk and used to activate WWW links.

This example shows how a conventional paper document can be used as the key providing access to the full range of electronic multi-media on the World-Wide Web.

Conclusions

In this paper we have described the use of animated paper documents to provide paper interfaces to the World-Wide Web. This combines the power of electronic hypertext and the convenience of printed documents.

Electronic publishing is a rapidly growing area with tens of thousands of titles in print on CD-ROM and hundreds of new titles being published each week. Direct publication exclusively in electronic form on the Internet is also growing. However, the problems of screen-based publishing - poor readability, limited view, slow access, inability to add personal annotations and so on - have limited its use to specialised applications. We believe that computer additions to printed texts offer a more promising approach, especially when delivered over a communications network.

BEST COPY AVAILABLE

Current work on animated paper documents is investigating both the underlying technology of the DigitalDesk and also new applications of mixed-media publication for educational material and more general use. We are particularly interested in using printed documents as the key to the delivery of electronic documents via network computers.

References

[Arnold & Gosling 1996] Ken Arnold & James Gosling (1996): *The Java programming language*, Addison-Wesley, ISBN 0-201-63455-4.

[Berners-Lee et al. 1994] Tim Berners-Lee, Robert Cailliau, Ari Luotonen, Henrik Frystyk Nielsen & Arthur Secret (1994): The World-Wide Web, *Communications of the ACM* 37(8), 76-82.

[Birrell et al. 1993] Andrew Birrell, Greg Nelson, Susan Owicki & Edward Wobber (1993): Network objects, *14th ACM Symposium on Operating System Principles*, December 1993, 217-230.

[Brown & Meehan 1993] Marc Brown & James Meehan (1993): *The FormsVBT reference manual*, DEC Systems Research Centre.

[Brown & Najork 1996] Marc Brown & Marc Najork (1996): *Distributed active objects*, DEC Systems Research Centre Report 141a; published in *Computer Networks and ISDN Systems* 28, 1037-1052.

[Harding & Quinney 1990] Robert Harding & Douglas Quinney (1990): Computer illustrated texts for teaching numerical analysis, *Computers and Education* 15(1-3), 205-211.

[Harding et al. 1997] Robert Harding, Steve Lay, Peter Robinson & Richard Watts (1997): New technology for interactive computer aided learning, *3rd Annual conference on alternative learning technology*, Glasgow, *Journal of the Association for Learning Technology* 5(1), 6-12.

[Manasse & Nelson 1991] Mark Manasse & Greg Nelson (1991): *Trestle reference manual*, DEC Systems Research Centre Report 68.

[Nelson 1991] Greg Nelson (1991): *Systems programming with Modula-3*, Prentice-Hall, ISBN 0-13-590464-1.

[Robinson 1995] Peter Robinson (1995): Virtual offices, *Royal Society discussion meeting on Virtual reality in society, science and engineering*, BT Publication SRD/R5/1.

[Robinson et al. 1997a] Peter Robinson, Dan Sheppard, Richard Watts, Robert Harding & Steve Lay (1997): Animated paper documents, *7th International Conference on Human-Computer Interaction*, HCI '97, San Francisco.

[Robinson et al. 1997b] Peter Robinson, Dan Sheppard, Richard Watts, Robert Harding & Steve Lay (1997): A framework for interacting with paper, *EuroGraphics '97*, Budapest, *Computer Graphics Forum* 16(3).

[Stafford-Fraser 1996] Quentin Stafford-Fraser (1996): *Video augmented environments*, PhD dissertation, University of Cambridge Computer Laboratory.

[Stafford-Fraser & Robinson 1996] Quentin Stafford-Fraser & Peter Robinson (1996): BrightBoard - a video augmented environment, *ACM Conference on Human Factors in Computing Systems*, Vancouver.

[Wellner 1993] Pierre Wellner (1993): Interacting with paper on the DigitalDesk, *Communications of the ACM* 36(7).

[Wellner 1994] Pierre Wellner (1994): *Interacting with paper on the DigitalDesk*, Technical Report 330 (PhD dissertation), University of Cambridge Computer Laboratory.

Acknowledgements

The original DigitalDesk was built by Pierre Wellner, a research student in the Computer Laboratory at the University of Cambridge sponsored by Rank Xerox. Current work on animated paper documents is sponsored by the EPSRC under grant GR/J65969.

FREE SPEECH on the INTERNET: LEGAL, SOCIAL, and POLITICAL ISSUES

Richard S. Rosenberg
Department of Computer Science
The University of British Columbia
Vancouver, B.C. Canada
rosen@cs.ubc.ca

Abstract: This paper is a brief outline of some issues and concerns underlying the widespread opinion held in many countries that there are some serious problems that compromise the usefulness and promise of the Internet and that challenge the limited jurisdiction of the international community. It is claimed here that in the context of the vast amounts of information that daily flood the world's communications systems that carry Internet traffic, the objectionable portion is very tiny. [my subjective and informal view of course] In this paper, we will argue that content regulation of the Internet (and the much heralded Information Highway) by governments is both impractical and unnecessary.

Introduction

This paper is a brief outline of some issues and concerns underlying the widespread opinion held in many countries that there are some serious problems that compromise the usefulness and promise of the Internet and that challenge the limited jurisdiction of the international community. It should be emphasized that much of the concern derives from a series of highly publicized incidents that portray the Internet as a useful tool of pornographers, child molesters, bomb makers, and other criminals. The open and unfettered exchange of information surely includes some that is objectionable, some that is hateful, some that is legally obscene in many but not all counties, and some that is directed towards child pornographers and molesters. But in the context of the vast amounts of information that daily flood the world's communications systems, that carry Internet traffic, the objectionable portion is very tiny. [my subjective and informal view, of course] The world's attention need not be distracted from serious problems to marginal ones.

Although the issue of free speech on the Internet has been associated with the well-publicized public concern with obscenity and more particularly child pornography, many other areas of controversy have emerged. Consider the following issues, some of which will be addressed in more detail in this paper:

1. Pornographic pictures and stories (binaries), which may be obscene.
2. Sale and distribution of child pornography.
3. Offensive newsgroup postings and Web sites: sexual, racial, ethnic.
4. Violation of court-ordered publication bans, e.g. Karla Homulka related issues in Canada.
5. Threats of assault and violence.
6. Bomb manufacturing information.
7. Character assassination (libel).
8. Release of confidential information.
9. Sexually explicit conversations in online chat rooms.
10. The Internet as a medium for the seduction of minors.

There are no accurate numbers to indicate how prevalent occurrences of such events are, even if they could be sufficiently well-defined to permit detailed calculations. The well-publicized paper from Carnegie-Mellon, popularly known as the "Rimm study" [Rimm 95] reported by [Elmer-DeWitt 1995] did a great disservice in this regard by claiming that the Internet was rife with pornography, a charge successfully challenged by [Hoffman and Novak 95] and many others.

In this paper, we will argue that content regulation of the Internet (and the much heralded Information Highway) by governments is both impractical and unnecessary. It is impractical because the global network presents extraordinary problems - legal, political, and technical - to those governments seeking to regulate it. And it is unnecessary because for the most part, excluding certain universally agreed upon crimes for which international agreements currently exist, the community of Internet users can determine appropriate, consensual procedures for acceptable behavior.

Background Information

Pornography does not equal obscenity, in North America at least. Pornography is generally available in hard copy, in movies, on television, and of course on computer bulletin boards, the Internet and Web sites. Obscenity is a legal term and although the definitional criteria may vary from country to country, its production and sale is illegal, whether done electronically, over networks or in print and in film. In the U.S., the Miller test is the current legal definition of obscenity. [Hawkins and Zimring 88] The first requirement of this three-part test refers to "contemporary community standards," an increasingly muddied concept in the age of the Internet. What makes obscenity difficult to define is that different cultures, countries, and indeed people have quite different thresholds, including those required to make the legal decisions, the judges.

Child pornography is universally abhorred and laws exist in most countries prohibiting its manufacture, sale, distribution, and indeed ownership. Dealing with it is one of the motivating forces underlying tentative attempts for international cooperation in regulating content on the Internet. Note that in the U.S., a new law was passed updating the child pornography laws. It is of serious concern to free speech advocates because it now includes computer-generated pictures that depict underage children engaging in sexual acts, even though no real world children were ever involved. Given recent events in Belgium, it is no surprise that its government is concerned about the seduction of children, but why is the Internet believed to be a prominent vehicle. For more information about a host of legal issues associated with the Internet, see [Rosenberg 97a].

Some Significant Examples of Offensive Internet Content

The following examples are intended to suggest the range of issues that have prompted concern about the nature of some of the content carried on the Internet. The examples were chosen to be suggestive but clearly not exhaustive; many others are available but the point is to acknowledge that Internet content certainly extends from the innocuous to the illegal and in this regard mirrors that available on the more traditional media.

Pornography (Images)

The cover story of Time magazine of June 3, 1995 seemed to say it all: "Cyberporn, Exclusive: new study shows how pervasive and wild it really is. Can we protect kids - and free speech?" [Elmer-DeWitt 95] This article has had an enormous impact in sensationalizing the Internet as being rife with pornography, obscenity and worse and laid the groundwork for support for regulatory Internet legislation. Its main results were challenged but the Internet had become synonymous with sexually offensive material and a clear danger to the American home. Such was the background that contributed to the overwhelming approval of the Communications Decency Act of 1996 by the Congress.

Pornography (Text)

Jake Baker was a University of Michigan student who wrote very violent, sexually explicit stories and circulated them on the Internet. He was arrested in February 1995 for supposedly including a female classmate, in one of his stories, who is tortured and murdered. In addition, he apparently exchanged e-mail with a friend in Canada, in which they discussed committing an actual murder involving rape and torture. Mr. Baker's case was subsequently dismissed on June 21, 1995 because no real threats or conspiracy could be proven. [Godwin 1995]

Racism (U.S.)

Dan Gannon had a mission: To post as many anti-holocaust messages as possible and to as many newsgroups as possible. For years, Gannon had been providing a forum for revisionists to spout their line that the holocaust did not happen. On March 10, 1994 Gannon posted a special diatribe (on the alt.censorship newsgroup) on the occasion of a perceived restriction in his otherwise open access to the Internet. Part of this plea follows: "Administrators at Netcom have told me I cannot post any more messages about Holocaust Revisionism to any newsgroups except certain newsgroups they have specified. I have no choice but to comply. There is a lobby which opposes any critical examination or questioning of the 'Holocaust story' or of Israeli policy. . . Following are the ONLY newsgroups Netcom says I am still allowed to post to: alt.revisionism, talk.politics.misc, soc.culture.german, soc.culture.jewish, soc.rights.human, alt.discrimination, alt.conspiricy, alt.illuminati, alt.individualism, alt.mindcontrol, alt.politics.correct, alt.politics.reform, and alt.censorship." [Gannon 94] This story is an example of the Internet community policing itself, not to censor but to respond to a serious violation of Internet etiquette and responsible behaviour, the irresponsible consumption of bandwidth to the detriment of other users.

Racism (Canada)

"In an unprecedented move, the Canadian Human Rights Commission has ordered hearings into complaints that Holocaust denier Ernst Zundel is promoting hatred on the Internet. Commission chief Max Yalden said yesterday that he believes that the commission has jurisdiction to shut down Mr. Zundel's Web site, even though it's based at a computer in California." [Bueckert 96] The legality of the Canadian view will certainly be challenged. In the U.S. hateful and extreme anti-semitic and racist expression is protected by the First Amendment. But note that in Germany, denial of the existence of the Holocaust is against the law and not protected. How will such dramatically opposing views and indeed laws be accommodated by international agreements?

Other Examples

There are many other examples including restrictions of access to certain newsgroups carried by CompuServe in Germany, the murder of a women apparently by a man who first made contact with her over the Internet (a somewhat dubious reason to control content on the Internet, given that the telephone could similarly be accused of contributing to orders of magnitudes greater numbers of similar crimes), the concern by Quebec and France about the prevalence of English as the language of choice on the Internet, the increasing growth of junk mail and spamming as an interference with normal traffic, and finally, politically motivated restrictions on Internet access by such countries as China, Singapore, Iraq, and Iran.

For many, the Internet has created a global community, offering possibilities for new forms of cooperation and information-sharing, as well as for exerting political and economic pressure. However, for others it has been defined by a host of problems that must be solved before the economic potential can be realized.

Possible Approaches to the Control of Internet Content

The following list is a mixture of legislative, voluntary, and technical approaches. With the present space limitations, very little detail is included, but enough it is hoped to foster discussion and debate. More detail is provided, however, in [Rosenberg 97b].

Use of Existing Laws

There are existing laws against obscene material and if such material is circulated on the Internet, it may be subject to these laws. However, as noted laws vary from country to country and the most restrictive laws would not be acceptable in the more liberal countries. Restrictions on political speech will certainly be opposed by most Western countries, that have already criticized Singapore and China, for example, for their anti-democratic actions.

New Government Legislation

The U.S.A. did pass legislation with respect to "indecent material" on the Internet, the Communications Decency Act of 1996, part of the Telecommunications Act. Its constitutionality was immediately challenged by many civil liberties groups as well as software and hardware companies and publishers, broadcasters, and others. The U.S. Supreme Court after hearing arguments in March, upheld the lower court decision in June 1997, thereby declaring certain sections of the Act unconstitutional. The issues being discussed will of course have relevance for the present concern. In addition, thirty or so State laws have already been passed or are in the works in the U.S. Other countries may also be interested in adopting restrictive legislation. In Canada, the

Information Highway Advisory Council's 1995 Final Report [IHAC 95] and the Government's 1996 response [Industry Canada 96] speak of legislation as a possible necessary approach, but also stress public education, voluntary measures, and international agreements.

Blocking or Filtering Programs

Such programs as CyberSitter, NetNanny, and Surfwatch provide a practical means for parents to restrict access to Web sites and newsgroups, either by name or by label. They were cited by the Pennsylvania court, as well as the Supreme Court, that ruled that certain parts of the Communications Decency Act were unconstitutional. However, they are not an unmixed blessing. Although certain ones permit parents to set the parameters, others screen on the basis of assumptions that may not be acceptable or even apparent. [McCullagh 96]

Direct Parental Control

Why not let parents assume direct responsibility for their children's viewing behavior. It is not always easy or convenient but given that parents do control their children's behavior in other contexts - television and movie viewing, curfews, restrictions on travel, and warnings with respect to strangers - it is not unreasonable to consider Internet activity as yet another aspect of parental responsibility.

Trust in Responsible Behavior

In previous papers, Rosenberg, [Rosenberg 93] and [Rosenberg 95], argued that with respect to the viewing of sexual material on public workstations at universities, libraries, and community centers, the following set of principles may serve as a way to respond to genuine concerns:

Administrative Principles

(1) Do not treat electronic media different from print media, or traditional bulletin boards merely because they can be more easily controlled.
(2) Do not censor potentially offensive material on networks; encourage the use of sexual harassment procedures, if appropriate.
(3) Be aware of your responsibility with respect to the uses and misuses of your facilities. However, do not use cost of services as an excuse to censor and limit access.
(4) Trust and educate people to be responsible.

Social Principles

(1) Issues will proliferate beyond the ability of organizations to control them by rigid policies.
(2) Occasional offensive postings do not detract from the benefits of electronic networks.

Self-regulation by the Internet Community

There is a vocal segment of Internet users, that argues that the Internet as a new and powerful medium, essentially created by its users, owes no allegiance to existing governments and indeed has the right and obligation to create its own operating procedures. Witness the opening paragraph of a manifesto released by John Perry Barlow, a cofounder of the Electronic Freedom Frontier: "Governments of the Industrial World, you weary giants of flesh and steel, I come from Cyberspace, the new home of Mind. On behalf of the future, I ask you of the past to leave us alone. You are not welcome among us. You have no sovereignty where we gather." [Barlow 96] Of course, the Internet is not a separate government and its users are not free to act independently of existing laws, in spite of the fact that detection and enforcement may be difficult or even impossible.

Barriers to Effective Regulation

The following discussion is of necessity abbreviated, all of the sections deserving considerably more elaboration. The purpose here is to be provocative and to argue that difficulties and barriers will be constant companions of attempts to control the evolving technology.

Anonymity

Dealing with objectionable or offensive material will necessarily involve confronting the use of anonymous remailers. The treatment of anonymity has differed and will continue to differ from country to country. A proper treatment is beyond the scope of this paper.

Cryptography

Encrypting obscene material prior to transmission is a way to hide it from prying eyes and so international cooperation will be necessary to prevent this occurrence. The resistance of the Internet community to governments controlling strong encryption is well known and will present enormous enforcement difficulties. However, there is also considerable pressure from the business community to develop acceptable security procedures to encourage the growth of commercial activity on the Internet. Effective and convenient cryptography standards are the goal, the wishes of the Internet community notwithstanding.

Intellectual Property Rights

The enforcement of copyright laws would limit the amount of binaries in circulation on the Internet, given that most of the images are either scanned in from magazines and videos or downloaded from electronic bulletin boards. But such enforcement is difficult and expensive.

Jurisdictional Issues

The Internet is a worldwide phenomenon. The implications of this fact are far reaching. No one country can force others to accept its particular view of the world, or more importantly its legal system. The call for international agreements to regulate the Internet with respect to an increasing number of perceived problems are either naive or purposefully misleading.

Legal Responsibility of ISPs (Internet Service Providers)

In many countries, the legal responsibilities of ISPs are unclear. In some cases, they have adopted voluntary codes of behavior that require them to remove access to certain newsgroups and Web sites , when informed that illegal material may be available. The obvious problem is that they are then assuming the role of censor, without public consensus. In some countries, governments are moving to pass legislation to regulate the activities of ISPs. Last year the government of Singapore did pass such legislation. [HRW 96] Whereas the responsibility of ISPs to guarantee the privacy rights of their users must be enforced, they should not assume the role of censors, especially not to forestall potential government regulation.

Technical Issues

Could the Internet be effectively regulated? A global distributed system, originally designed to withstand a nuclear attack is very resistant to disruption or control. The numbers of sophisticated users throughout the world is very large and if there is one principle that is universally adhered to, it is free and open expression. A restricted Internet will be fought vigorously and most likely effectively. Of course, most users around the world would probably, in the end, obey restrictive laws.

Conclusions

Consider the following comments taken from the unanimous decision of the three-judge Pennsylvania panel in finding that certain provisions of the Communications Decency Act of 1996 violate the First Amendment of the U.S. Constitution, with respect to the limitations on free speech:

> **District Judge Buckwalter**:
> The thrust of the Government's argument is that the court should trust prosecutors to prosecute only a small segment of those speakers subject to the CDA's restrictions, and whose works would reasonably be considered "patently offensive" in every community. Such unfettered discretion to prosecutors, however, is precisely what due process does not allow. [EPIC 96]
>
> **District Judge Dalzell**:
> Cutting through the acronyms and argot that littered the hearing testimony, the Internet may fairly be regarded as a never-ending worldwide conversation. The Government may not, through the CDA, interrupt that conversation. As the most participatory form of mass speech yet developed, the Internet deserves the highest protection from governmental intrusion. Just as the strength of the Internet is

> chaos, so the strength of our liberty depends upon the chaos and cacophony of the unfettered speech the First Amendment protects. [EPIC 96]

There are legitimate concerns about content on the Internet, but government enforced regulation is not the best way to deal with them. Difficulties and alternatives have been presented but only an informed and sufficiently motivated public can make a difference. In discussing the nature of a liberal democracy, the Canadian political scientist C.B. Macpherson analyzed the opinions of the American scholar, John Dewey as follows:

> "He has few illusions about the actual democratic system, or about the democratic quality of a society dominated by motives of individual and corporate gain. The root difficulty lay not in any defects in the machinery of government but in the fact that the democratic public was "still largely inchoate and unorganized," and unable to see what forces of economic and technological organizations it was up against. There was no tinkering with the political machinery: the prior problem was 'that of discovering the means by which a scattered, mobile, and manifold public may so recognize itself as to define and express its interests.' The public's present incompetence to do this was traced to its failure to understand the technological and scientific forces which had made it so helpless." [MacPherson 80]

References

[Barlow 96] Barlow, J.P. (1996). Declaration of the independence of cyberspace. February 8. Accessed from the Web page with the URL: <http://numedia.tddc.net/scott/declaration.html> on July 15, 1997.

[Bueckert 96] Bueckert, D. (1996). Human Rights panel to probe Zundel's Web site. Canadian Press, November 22.

[Elmer-Dewitt 95] Elmer-DeWitt, P. (1995). Cyberporn, *Time*, July 3. 32 - 39.

[EPIC 96] Electronic Privacy Information Center (1996). ACLU v. Reno, highlights and analysis. June 12. Accessed from the Web page with URL: <http://www.access.digex.net/~epic/cda/highlights.html> on January 27, 1997.

[Gannon 94] Gannon, D. (1994). Locked out of Usenet newsgroups. As posted on alt.censorship on March 10.

[Godwin 95] Godwin, M. (1995). Baker Column, *Internet World,* June. Accessed from the Web page with URL: <http://www-swiss.ai.mit.edu/6095/articles/computer-crime/godwin-baker-article-june95.html> on July 8, 1996.

[Hawkins and Zimring] Hawkins, G. & Zimring, F.E. (1988). *Pornography in a Free Society*. New York: Cambridge University Press, p. 130.

[Hoffman and Novak 95] Hoffman, D. L. & Novak, T. P. (1995). A detailed analysis of the conceptual, logical, and methodological flaws in the article: "Marketing pornography on the Information Superhighway." Accessed from the Web page with URL: <http://www.2000.ogsm.vanderbilt.edu/rimm.cgi> on July 30, 1995.

[HRW 96] Human Rights Watch (1996) Letter to George Yeo, Singapore Minister for Information and The Arts. Available on the Web page with URL: < gopher://gopher.igc.apc.org:5000/00/int/hrw/asia/asia/4>.

[Industry Canada 96] Industry Canada (1996). Building the Information Society: Moving Canada into the 21st Century. May. Available at the Web page with URL: <http://strategis.ic.gc.ca/cgi-bin/dec/wwwfetch?/sgml/ih01103e_pr702.sgml>.

[IHAC 95] Information Highway Advisory Council, Industry Canada (1995). The Challenge of the Information Highway (Final Report) September. Available at the Web page with URL: <http://strategis.ic.gc.ca/cgi-bin/dec/wwwfetch?/sgml/ih01015e_pr702.sgml#engdoc/>.

[Lewis 95] Lewis, P.H. (1995). An Internet author of sexually violent fiction faces charges. *The New York Times* , February 11.

[MacPherson 80] MacPherson, C.B. (1980). *The life and times of liberal democracy* . Oxford, England: Oxford University Press., p. 73.

[McCullagh 96] McCullagh, D. (1996). The CyberSitter diaper change. *The Netly News*, December 20. Accessed from the Web site with URL: <http://netlynews.com/> on January 21, 1997.

[Rimm 95] Rimm, M.(1995). Marketing pornography on the Information Superhighway. Accessed from the Web page with URL: <http://trfn.pgh.pa.us/guest/mrstudy.html> on July 11, 1995.

[Rosenberg 93] Rosenberg, R.S. (1993). Free speech, pornography, sexual harassment, and electronic networks. *The Information Society,* 9 (4), 285-331.

[Rosenberg 95] Rosenberg, R.S. (1995). Free, speech, pornography, sexual harassment, and electronic networks: an update and extension. In Mackaay, E., Poulin, D. & Trudel, P. (eds.), *The Electronic Superhighway*. The Hague, The Netherlands: Kluwer Law International, 127-158.

[Rosenberg 97a] Rosenberg, R.S. (1997). *The Social Impact of Computers*. San Diego, CA: Academic Press.

[Rosenberg 97b] Rosenberg, R.S. (1997). Free speech on the Internet: controversy and control. *The Ethics of Electronic Information in the 21st Century,* Memphis, TN. (To appear)

Acknowledgments

To my colleagues in Electronic Frontier Canada, David Jones and Jeff Shallit and my friend, and president of the B.C. Civil Liberties Association, Kay Stockholder and to NSERC for financial support.

Web-based Course Delivery and Administration using Scheme

Filippo A. Salustri
Department of Industrial and Manufacturing Systems Engineering
The University of Windsor
Windsor, Canada
e-mail: salustri@uwindsor.ca

Abstract: This paper discusses the use of a small Web-based tool for course delivery and administration called HAL (HTML-based Administrative Lackey), written in the Scheme programming language. Various problems encountered in administering the course are discussed and proposed solutions are presented. The design and implementation of HAL is then briefly described, with examples. Based on the apparent success of HAL, it appears that though larger and more complex course delivery systems are suitable in some circumstances, there are situations in which simpler and smaller systems are better suited.

Introduction

Administering and delivering course material to large groups of students can be very difficult. This is especially true in courses which significant technical content or requiring laboratory or other group efforts by the students. Typical problems include: organizing student groups, scheduling tutorials or laboratories taking into account possibilities of course conflicts, availability of teaching assistants and other resources, updating and disseminating information to students without tremendous waste of paper, controlling student evaluation, managing and tabulating student marks/grades, and ensuring that students who may not know one another can communicate with each other remotely (over a computer network, for example).

Clearly, there is a role that the World Wide Web can play to resolve these issues effectively. This paper discusses a tool being developed by the author to provide Web-based services for a large first-year undergraduate course in computer-aided design (CAD).

The Problem

The CAD course is a required course for all first-year undergraduate engineering students at the University of Windsor (currently, enrollment in the course is over 200). The goal of the course is to give the students some exposure to group-based engineering design techniques, especially in terms of concepts of *concurrent engineering* and *total design*. The following problems with administering the course and delivering material have been identified:

1. Students are required to work in small groups to actually design a product (this year, the product is a movable garden hose carrier). It is important to establish groups in some manner that hopefully maximizes the educational experience. A simple random selection process was not considered sufficient. Furthermore, since most first-year students are registered in the "General Engineering" program, it was impossible to use the chosen fields (mechanical engineering, electrical engineering, etc.) as a basis for group formation.
2. The University Registrar's Office is unable to provide an electronic list of student enrollment that is current. Students are allowed to drop a course - even a "compulsory" one such as CAD - up to two months after the start of the term. As enrollment changes, groups may need to be rearranged. Group management becomes a semester-long task.

3. Students must be able to communicate with each other effectively using electronic mail. Every undergraduate student has a computer account for the duration of his/her program. Group-based mailing lists were an obvious mechanism to expedite communications, but the administrative load of expecting the University to create and administer such lists on a semester-by-semester basis was unacceptable.
4. Disseminating information to such a large group of students is a difficult task. Photocopying alone can significantly drain already thinly stretched teaching resources, and can involve a tremendous waste of paper. Getting information to students in a *timely* manner also becomes problematic when using paper.
5. Six teaching assistants are assigned to the CAD course, one for each laboratory/tutorial session. Tutors are charged with supervising their students, and marking progress reports for individual students and groups. The marks must be tabulated and stored consistently to simplify the assignment of final grades at the end of the semester. No standardized way of doing this currently exists in the department.

Formulating a Solution

Obviously, given the nature of the course, there is great potential for *Web-based* course delivery and administration in this situation. Two existing systems, [WebCT] and [Virtual U], were investigated. Both systems are quite large and provide many facilities besides those required for the CAD course. The administrative overhead associated with acquiring, installing, and maintaining the software, as well as designing the courseware itself, is considerably more than is reasonable at this time in the author's faculty.

The author also has an interest in developing Web-based administrative tools for educational and administrative environments. Most of the author's current research involves the Scheme programming language; it was considered desirable to be able to integrate the delivery system for the CAD course with the author's existing work.

In light of these issues, it was decided that a smaller, simpler solution should be tried. The system to be developed would be designed to meet the above-mentioned requirements, and possible to have the flexibility to be extended in the future, should there be sufficient interest. The proposed solution includes the following components.

Formation of student design groups. The use of so-called "personality tests" has found some popularity in industrial settings. There is also research to suggest that student design groups set up to have diversity of personalities have performed better than groups established by other means [Wilde 93]. A small personality test was available to the author; the test is a component of the delivery software. All registered students are required to take the test; the results are presented to them in a form they can understand. The results, along with course schedule information, are also used to assign students to groups. A simple algorithm was devised that automated the assignment process. A student would simply take the personality test, and as a result be assigned to a particular design group and laboratory/tutorial session. As enrollment changes, individuals can be automatically re-assigned to other groups as required.

Maintaining registration information. Since information from the University Registrar's Office was useless, the delivery system would have to allow students to "register" for the course. This allows enrollment information to be gathered quickly, and in a form immediately useful by the system.

Facilitating communications. The registration component of the delivery system requires students to supply the login of their University computer account, which is identical to their local e-mail address. This information, combined with the design groups database, is used to provide students with the means to e-mail messages to each other by name rather than login (many students do not know each other's login names), as well as sending messages to every member of their group without necessarily knowing the others' e-mail addresses. This is intended to allow groups to communicate outside of the assigned tutorial sessions.

Disseminating information. The system can alert students, by group or individually, that some information on the Web pages for the course has been updated. This allows the dissemination of information more efficiently, without tremendous waste of paper, time, etc. and with a higher degree of assurance that each student will actually be aware that updates have occurred. All course information except for lecture material is kept in Web pages that each student can access via the delivery system.

Assigning and recording grades. Since every student registers for the course via the delivery system itself, there is ample information to develop a unified internal database able to maintain grades assigned by the teaching assistants. Teaching assistants are recognized by the system, and are allowed to use it to enter grades. The system automatically tabulates semester and final grades based on this input.

Implementing the Solution

The course delivery system is implemented as a single CGI program able to return whatever HTML pages and forms are required. The program is able to develop pages dynamically, in response to user input, and to carry out other computations, such as assigning students to groups, and maintaining various databases. The Scheme programming language was used to implement the CGI.

The Scheme Programming Language

Scheme [IEEE 90] is a formalized dialect of Lisp. The particular implementation used, SCM by Aubrey Jaffer [Jaffer], supports the base language as well as POSIX-compliant extensions for the manipulation of files, and has a large library of portable packages. SCM programs can also be called as batch files, which allows it to be used for CGIs.

Scheme was used because it is a very simple yet powerful language, it has a very small "footprint" (significantly smaller than Perl), and is very efficient. It is also the language used by the author for a number of other research efforts.

The Scheme CGI

The first step in implementing the system was to develop a library of Scheme functions to facilitate creating HTML pages. The ultimate result was more than just able to decode CGI query strings, etc. The Scheme CGI in fact implements functions that have direct equivalents in HTML. While this might seem to simply duplicate what HTML already provides, the approach allows a far higher degree of integration of HTML and Scheme. A single consistent syntax allows programmers to develop Scheme programs that can also create and query arbitrary HTML pages, without concern for the syntax of HTML itself.

For example, without the Scheme CGI library, one might define the following function to display on the standard output a list of test string arguments as a compact unordered list in HTML, one string per HTML list item.

```
(display "<UL COMPACT>") (newline)
(for-each (lambda (string)
            (display "<LI>")
            (display string)
            (display "</LI>")
            (newline))
          a-list-of-strings)
(display "</UL>") (newline)
```

Figure 1: Plain Scheme code for an HTML unordered list.

The Scheme CGI library, however, implements various functions that allow such lists to be nested, or contain other HTML constructs, which the above function does not do. For example:

```
(unordered-list
  (!compact)
  (item "This is the" (bold "first") "string.")
  (item (unordered-list
          (item "The first sub-item.")
          (item "The" (italics "second") "sub-item.")))
  (item "The third string."))
```

Figure 2: An HTML unordered list using the Scheme CGI library.

The CGI query string is decoded as a list of name/value pairs in a Scheme variable called `*qargv*`, and values can be searched for by name with the function `getarg`. All the environment variables passed to a CGI are available as well (e.g. the function `(server-software)` returns the value of the `SERVER_SOFTWARE` environment variable. All the structures in HTML 3.2 are represented by Scheme functions. Attribute flags, like "`!compact`" above, are by convention named with an initial exclamation mark; attributes that take such values, such as "`:name`" start with a colon and take a single argument that can be a text string or another Scheme CGI library item.

Constructing an HTML page is quite simple. The following example shows how the initial login page is constructed for the CAD course system.

```
(html (head (title "85-131: Computer-Aided Design"))
      (body (:color/bg "black")
            (:color/text "white")
            (heading 1 (hal-logo) "Access Verification Page")
            (paragraph "Hi.  I'm HAL, the HMTL-based Administrative
                        Lackey, for this course."
                        :breakline
                        "To continue, you need to enter your"
                        (italics " login") " and your"
                        (italics " password.")
                        "When you're done, hit the"
                        (bold " submit") " button.")
            (form (:action "/cgi-bin/fil/85-131")
                  (:method "get")
                  (input "hidden" (:name "action") (:value the-action))
                  :breakline
                  (bold "Enter Login here:")
                  (input "text" (:name "login") (:size 10))
                  :breakline
                  (bold "Enter Password here:")
                  (input "password" (:name "pw") (:size 10))
                  :breakline
                  (input "submit"))))
```

Figure 3: Construction of a whole HTML page using the Scheme CGI library.

This will return a structure containing all the HTML fragments needed to product a Web page. The function "`/show-page`" is used to actually output this structure on the standard output.

The complete library is quite small, less than 50KB of commented Scheme code, and runs at speeds at least comparable to equivalent Perl code.

BEST COPY AVAILABLE

The HTML-based Administrative Lackey

The HTML-based administrative lackey (or "HAL") is a Scheme program that uses the Scheme CGI to centralize all the services needed for a given course, in one, integrated unit. The program simply returns an HTML page corresponding to the requested query. In the progress of creating the pages, the program may load various other files or perform other calculations.

The first time a user accesses HAL, by selecting the appropriate link on the web page of the CAD course, the user is asked to log in. Only registered students (and teaching assistants and instructors) are allowed passed this point. HAL searches for a student file by the login/password combination. The student file contains information such as the student's name, personality type, design group number and a list of the student's marks. If a student who is not registered attempts to log in, the student is presented with a registration page asking for the student's full name. This page also allows the student to take the personality test. Once this information is provided and the test is taken, HAL assigns the student to a design group, updates all relevant databases, and presents the student with the "main" page. This page gives the student the options of viewing various other pages relevant to the course, advices the student regarding recent updates and announcements, and provides the means for the student to send messages to the other members of his/her group, a teaching assistant, or the instructor. The student may also view his/her record, including grades, in a read-only format.

Finally, HAL can identify when it is running as a CGI and when it has been invoked from a simple command-line. In the latter case, the program will run only for the administrator and will allow various specialized tasks, such as identifying teaching assistants and instructors, initializing databases, etc.

Future Directions

Clearly, HAL is not intended to be as all-encompassing as are some of the other Web-based course delivery systems mentioned earlier. However, it is ample for the purposes of the CAD course taught by the author. If the HAL facility meets with success, the author will bring it to the attention of other instructors, in the hope that it will help them as well.

In this eventuality, it will be necessary to create a *meta*-HAL facility to assist other instructors in developing other HAL programs for their own courses. It is also possible that other administrative programs, similar to HAL, may be developed to assist in the daily administration of the author's department.

Conclusions

This paper has discussed the problems of administering a large class, and how a simple, small, and efficient language like Scheme can be used to quickly construct a system for aiding in the delivery of material and course administration. It appears that the HAL system will significantly improve the efficiency by which administrative details are dealt with, freeing the instructor and teaching assistants to focus their attention on helping the students learn. The larger systems that are available may be well suited in some circumstances, but clearly there are other situations for which a simpler, smaller system is advised.

References

[WebCT] World-Wide Web Course Tool, http://homebrew.cs.ubc.ca/webct/.

[Virtual U] The Virtual University, http://www.emediadesign.com/sfu/courses.html.

[Wilde 93] D. J. Wilde. (1993). Mathematical resolution of MBTI creativity data into personality type components, *ASME Design Theory and Methodology Conference*, T. K. Hight, ed. 37-43.

[IEEE 90] IEEE. (1990). *IEEE Standard for the Scheme programming language*. IEEE Std 1178-1990.

[Jaffer] Aubrey Jaffer's Meta Home Page, http://www-swiss.ai.mit.edu/~jaffer/index.html.

Law's Domain: Trademark Protection for Internet Address Names?

KURT M. SAUNDERS
Duquesne University
School of Law
U.S.A.
70752.1401@compuserve.com

Abstract: Internet domain names, used in e-mail and website addresses, have become a business asset. They may incorporate the business' trademark or be registered as a trademark under United States law. Domain names are assigned on a first come, first served basis by InterNIC, an Internet address registration service. However, because there can be only one registered owner of a domain name, the potential for trademark infringement and dilution arises when one business registers a domain name that is confusingly similar to or tends to weaken the trademark of another business. The possibility of litigation to resolve such a trademark dispute suggest that a business act with prudence when establishing an Internet presence.

Introduction

Domain names function as the address system for the Internet, allowing consumers to identify and locate particular businesses in much the same way as a street address or telephone number. As the number of people who have access to the Internet grows exponentially each year, domain names are emerging as a valuable tool for businesses to attract and inform potential consumers about their products and services. Consequently, domain names are becoming an increasingly important business assets in much the same manner as trademarks. Indeed, some domain names may incorporate or function essentially as trademarks. In those instances, important questions arise about the extent of protection afforded by trademark law. Such questions will only increase as more and more businesses seek to create a presence on the Internet.

Selection and Registration of Domain Names

A domain name identifies the location and category of a particular website. The domain name "xyzyx.com," for example, contains a top-level and second-level name. There are seven top-level domain names: ".edu" for educational institutions; ".com" for commercial businesses; ".org" for noncommercial organizations; ".int" for noncommercial international organizations; ".net" for network gateways; ".gov" for governmental offices; and ".mil" for the military. Second-level names, such as "xyzyx," are more specific identifiers and exist within top-level domains.

The United States National Science Foundation (NSF) created the International Network Information Center (InterNIC) to serve as the central information source for the Internet. The NSF has contracted with Network Solutions, Inc. (NSI), a private firm, to assign and register domain names. To become accessible on the Internet, a business must have an Internet protocol address and obtain from NSI a domain name, which is then registered with InterNIC. Domain names are assigned on a first come-first served basis, and only one domain name per business may be registered. InterNIC does not investigate to determine if a requested domain name incorporates or violates a registered trademark; instead, the applicant has the responsiblity to do so.

Under current policy, applicants must confirm that: (1) the applicant's statements in the application are true and that the applicant has the right to use the domain name requested, (2) the applicant has a bona fide intention to regularly use the domain name, (3) the registration and use of the name does not interfere with or infringe the intellectual property right of any other party in any jurisdiction, and (4) the applicant is not seeking to use the name for any unlawful purpose, including unfair competition. The applicant must agree to indemnify NSI against all disputes arising from the use or registration of the domain name. NSI does not arbitrate or adjudicate trademark disputes involving domain names. Instead, when a dispute arises, NSI will put the domain name on "hold" pending the dispute's resolution. [Baum & Cumbow (1996)]

Trademark Law in the United States

A trademark is "any word, name, symbol, or device, or any combination thereof ... used ... to identify and distinguish ... goods ... from those manufactured or sold by others and to indicate the source of the goods." [15 United States Code § 1127] Trademarks are classified in order of increasing distinctiveness: generic, descriptive, suggestive, arbitrary, and fanciful. The more distinctive the mark, the greater the protection afforded by law. [see Dueker (1996)]

Generic marks, which define or are synonymous with particular products, are not protectable. A mark such as "Breakfast Cereal" adopted for a breakfast food made from grain and sold as a quantity of small bite-size units typically eaten hot or cold and contained in a box is considered generic. However, a mark such as "Kleenex" or "Xerox" may become generic if it becomes so associated with its class of products so as to lose its distinctiveness. To prevent this, businesses often police their marks by bringing infringement actions any time the mark is used in a generic manner.

Suggestive marks are considered distinctive because they engage the consumer's imagination in determining the product source. "Westlaw" is suggestive because it links the legal research software with it's source, West Publishing Company. Fanciful and arbitrary marks are considered inherently distinctive. "Made up" words like "Exxon" or "Electrolux" are examples of fanciful trademarks; "Apple Computer" or "Sun Microsystems" are examples of arbitrary trademarks.

Descriptive marks may become distinctive when they acquire a secondary meaning that allows them to represent a particular source. To determine secondary meaning, courts consider the: (1) length and manner of use; (2) nature and extent of advertising and promotion; (3) efforts made to promote conscious connection in the public's mind between the trademark and the business; and (4) extent to which the public actually identifies the mark and the product or service.

Infringement under the Lanham Trademark Act [15 United States Code § 1114] occurs when the use of a mark by one person creates a "likelihood of confusion" with the mark of another. To determine the likelihood of confusion, the courts will examine the: (1) similarity of the marks; (2) similarity of the goods or services; (3) character of the market; (4) strength of the mark; and (5) intent of the alleged infringer. Ultimately, the concern is with the effect of the similarity on actual or potential consumers.

In addition, the Trademark Dilution Act [15 United States Code § 1125] creates a federal claim for dilution of "famous" trademarks. Dilution occurs when, through the use of a similar or identical mark, a strong and well-known registered trademark is weakened as a means of identifying and distinguishing a particular good or service. Unlike infringement, dilution can occur independent of any competitive relationship between the parties or confusion as to source. The owner of a famous mark may seek an injunction against a person who "willfully intended to trade on the owner's reputation or cause dilution of the famous mark." [Ibid.]

Overlay of Domain Names and Trademark Rights

NSI registration, unlike trademark registration, does not vest any rights of ownership of the domain name since it can be revoked. A problem arises when one business' domain name contains words constituting another business' trademark. Although no law expressly prohibits this, there may be consumer confusion as to the origin of the product or service, which constitutes infringement under the federal Lanham Trademark Act. Another infringement problem arises when someone requests a domain name, intending it to be confusingly similar to another's product or service, with the hope that it can be sold at a profit.

Similarly, a number of issues emerge when assessing the usefulness of applying trademark law to domain names. For instance, does or should the initial registration of a domain name containing a trademark in itself constitute infringement or dilution? Trademark law permits registration of the identical names for noncompeting goods and services, but only one domain name using that same name can be registered by NSI. For instance, trademarks such as "Morton's" salt, "Morton's" steakhouse, and "Morton's" appliances can coexist, but there can be but one "mortons.com" domain name. Another difficulty is the lack of geographic and product or service differentiation in Internet addresses. For example, the domain name "aaa.com" could idenitify the on-line address of the American Automobile Association or the American Arbitration Association even though consumers could easily distinguish between the services provided by each.

As the courts begin to address these issues, they will inevitably attempt to draw parallels between domain names and existing categories of protectable trademarks. One analogy might be geographic location names since domain names function as the Internet address system; however, geographic location names generally are treated as addresses, which are merely descriptive of a particular locale. Trademark protection does not extend unless they have acquired secondary meaning. Another analogy might be television or radio broadcast station call letters, which can be registered and enforced as trademarks by the prior user. Likewise, there may be an analogy to "vanity" telephone numbers used as pseudonyms. These are usually protectable, though purely generic alphanumeric terms or phrases are not. [Burk 1995]

Trademark Disputes involving Domain Names

Early trademark disputes involving domain names were settled out of court. One such case was *MTV Networks v. Curry* [867 F. Supp. 202 (S.D.N.Y. 1994)], where Curry, while employed by MTV, created an entertainment information Internet site registered as "mtv.com". When he left MTV, Curry refused to surrender the domain name and MTV sued. The parties later settled the matter, with MTV regaining ownership of the domain name.

In a similar case, Princeton Review registered the domain name "kaplan.com" and established an Internet site under that name. Stanley Kaplan Review, Princeton Review's archcompetitor in the standardized test preparation market, sued and demanded that Princeton Review cease its use of the name. The suit was settled by arbitration, which led to Princeton Review's surrender of the domain name. Another case involved a *Wired* magazine writer, Joshua Quittner, who registered the name "mcdonalds.com". The dispute was resolved when McDonald's agreed to underwrite the purchase of computer equipment for a local grade school in exchange for the rights to the domain name. [see Quittner 1994]

More recently, in the case of *Intermatic Inc. v. Toeppen* [1996 U.S. Dist. LEXIS 14878 (N.D. Ill. 1996)]. Intermatic alleged that Toeppen infringed its trademark when he registered and used the domain name "intermatic.com". The federal court reserved the infringement claim for trial to determine if there was consumer confusion, but held that the federal Trademark Dilution Act and the Illinois Anti-Dilution Act had been violated. According to the court:

> Toeppen's conduct has caused dilution in at least two respects. First, Toeppen's registration of the intermatic.com domain name lessens the capacity of Intermatic to identify and distinguish its goods and services by means of the Internet. Intermatic is not currently free to use its mark as its domain name. ... Such conduct lessens the capacity of Intermatic to identify its goods to potential consumers who would expect to locate Intermatic on the Internet through the "intermatic.com" domain name. ... Second,

> Toeppen's conduct dilutes the Intermatic mark by using the Intermatic mark on its web page. ... Dilution of Intermatic's mark is likely to occur because the domain name appears on the web page and is included on every page that is printed from the web page. [Ibid.]

Likewise, in *ActMedia, Inc. v. Active Media International, Inc.* [1996 WL 399707 (N.D. Ill. 1996)], the same court permanently enjoined the defendant from using or infringing ActMedia's trademark through use of "actmedia.com" as its domain name under the Illinois Anti-Dilution Act. In *Hasbro, Inc. v. Internet Entertainment Group, Ltd.* [1996 U.S. Dist. LEXIS 11626 (W.D. Wash. 1996)], another federal court issued a preliminary injunction requiring the defendant to cease using the domain name "candyland.com" on its Internet site containing sexually-explicit material as it infringed and diluted Hasbro's "Candy Land" trademark.

Conclusion and Recommendations

Trademark rights are inextricably intertwined with geographic location. Infringement is limited by the actual and possible geographic market and the extent of advertising. In fact, the same mark may be separately owned in different countries. The Internet, however, greatly enlarges the scope of a possible market since the Internet is worldwide. At present, there exists no international trademark registration agency, though article 16(1) of the General Agreement on Trade and Tariffs on Trade-Related Aspects of Intellectual Property Rights provides for cross-recognition and protection of the trademarks registered by signatory countries. Specifically, article 16(1) states: "The owner of a registered mark shall have the exclusive right to prevent all third parties not having his consent from using in the course of trade identical or similar signs for goods and services which are identical or similar to those in respect of which the trademark is registered where such use would result in a likelihood of confusion." In addition, the Paris Convention for the Protection of Industrial Property, to which the United States adheres, requires its signatories to treat citizens of other signatories the same for purposes of registration, but does not require recognition of a foreign trademark. Likewise, the European Union has also created a separate European mark.

As with the many other unique issues of law and governance raised by the growth of the Internet, it is likely in the long term that the issue of trademark protection for domain names will have to be addressed by an international accord. Ultimately, some new category of intellectual property protection for Internet domain names may need to be created. The European Union, for instance, has created a special form of intellectual property protection for computer databases, which receive very limited copyright law protection in the United States. [see Nimmer (1992)] Similarly, the United States recently enacted the Semiconductor Chip Protection Act [17 United States Code § 901-914], which creates another sui generis type of protection.

As a first step toward legal protection in the United States, however, a business should consider registering its domain name with the U.S. Patent and Trademark Office (USPTO) in order to gain an exclusive right of use. The USPTO will register a domain name as a trademark, so long as it is actually used as a trademark and not solely as an Internet address. Mere use as an Internet address will not meet the requirement that the mark be distinctive. The USPTO has issued an official policy statement on the use and registration of domain names as trademarks, which can be found at its site located at <http://www.uspto.gov/web/uspto/info/domain.html>. Next, a business should register its trademark with InterNIC as a domain name. In doing so, a business should attempt to discover, by conducting a comprehensive trademark search, if any other business or person is currently using a potentially infringing or diluting domain name. If the business' intended market extends beyond the United States, the search should be international in scope. Domain names may be searched through InterNIC's site located at <http://rs.internic.net/rs-internic.html>. If so, the business should notify InterNIC of this and demand that the alleged infringer or diluter immediately cease using the name.

References

[Baum & Cumbow (1996)]. Baum, R. & Cumbow, R. (1996). First use: Key test in internet domain disputes. *National Law Journal*, February 12, 1996, at C17.

[Burk 1995]. Burk, D., (1995). Trademarks along the infobahn: A first look at the emerging law of cybermarks. *Richmond Journal of Law and Technology* 1(1).

[Dueker (1996)]. Dueker, K. Trademark law lost in cyberspace: Trademark protection for internet addresses. *Harvard Journal of Law & Technology* 9(483).

[Nimmer (1992)]. Nimmer, R. (1992). *The law of computer technology* ¶¶ 5.08-5.09.

[Quittner 1994]. Quittner, J., (1994). Billions registered: Right now, there are no rules to keep you from owning a bitchin' corporate name as your own internet address. *Wired*, October 1994, at 50.

WWW Tools for Accessing Botanical Collections

Erich R. Schneider[1], John J. Leggett[1], Richard K. Furuta[1], Hugh D. Wilson[2], Stephan L. Hatch[3]
Texas A&M University, College Station, Texas, USA
Email: {erich,leggett,furuta}@csdl.tamu.edu, wilson@bio.tamu.edu, s-hatch@tamu.edu

[1]Center for the Study of Digital Libraries
[2]Department of Biology
[3]Department of Rangeland Ecology and Management

Abstract: The Texas A&M Bioinformatics Working Group, in furthering its goal of developing Web tools for accessing botanical information, has developed the Herbarium Specimen Browser. This tool allows investigators to panoramically survey the tens of thousands of specimens in the database from the S.M. Tracy herbarium, a major collection of preserved plants. While some of its implementation details (particularly its use of a full-text retrieval system to store the database and its specialized mapping software) are of interest, it also exhibits some properties which other designers may find worth consideration: support for pattern discovery, use of regularity in link destinations and sources, and employment of Javascript as an interface simplification mechanism.

1. Introduction

Since its inception in mid-1995, the Texas A&M Bioinformatics Working Group has pursued two primary activities: digitizing the contents of the S. M. Tracy Herbarium (a collection of over 200,000 preserved plants with a particular focus on the grasses of Texas) and creating Web tools for botanists and botanically-interested nonspecialists, mainly enabling viewing of the geographic distributions of various plant groups. For most of its history, the working group pursued these threads separately; preparatory work was being done on developing a system to allow the rapid input of specimen information from the herbarium, and the Web developments were done using information gathered by external entities. However, the threads have recently come together; input of herbarium specimen data has progressed to the point where it has become feasible (indeed, imperative) to produce Web-based tools allowing group members and the world at large to access the herbarium's resources electronically.

This paper describes the initial result of the confluence of these activity streams: the Herbarium Specimen Browser (http://www.csdl.tamu.edu/FLORA/tracy/hsb.html). Section 2 provides some background about our working group and the botanical collections known as *herbaria*. Section 3 explains implementation details behind the Specimen Browser. Section 4 describes some desirable properties the Specimen Browser possesses, which reflect principles that other Web designers may wish to consider. Section 5 concludes with speculations about future work.

2. About the Working Group and herbarium collections

The Texas A&M Bioinformatics Working Group (http://www.csdl.tamu.edu/FLORA/tamuherb.htm) is an interdisciplinary endeavor with participants drawn from three groups on campus: botanists from the Department of Biology, specializing in botanical taxonomy; botanists with similar expertise from the Department of Rangeland Ecology and Management, affiliated with the S.M. Tracy Herbarium; and computer scientists, specializing in hypermedia systems, from the Center for the Study of Digital Libraries.

Our working group is fortunate in that even before our current collaboration several of the biologist participants had begun developing Web materials on their own, and were therefore proficient in the Web technologies of the time (HTML markup and the structuring of Web information spaces) as well as in the use of commercial database programs. Consequently, they have been able to maintain information structured according to botanical needs and to maintain and develop much of the group's Web infrastructure, leaving the computer scientists to develop the "advanced" Web features.

One of the group's long-term goals is the replication of the information in the S. M. Tracy Herbarium in electronic form. The herbarium, one of 2639 in the world, is a collection of plant specimens which have been pressed, dried, and glued to cardstock sheets. Each specimen sheet has a label containing information on the collector, the location of collection, an accession number (a number uniquely identifying the specimen within the collection), and an identification of the specimen via a Latin scientific name, along with an indication of the taxonomist responsible for associate that name with that species. The process of assigning scientific names to plant species is one fraught with dispute and subject to continual evolution; as a result, many specimen sheets have annotations reflecting re-identification by later investigators.

The specimens in herbaria are vital to the practice of systematic botany, the branch of the field dealing with taxonomy. Herbarium specimens form the foundation of plant nomenclature, in that all scientific names (and the procedures for assigning them) are ultimately linked to specific *type specimens*. Also, herbarium collections are important in the construction of floristic manuals or *floras*. A flora is an exhaustive list, for a given region, of a given group of plants (e.g. of all grasses, or all flowering plants), their distributions within that region, and other information about them. A flora is deemed to possess greater veracity when the distributions in it are documented by herbarium specimens in addition to field observations. The over 1 million specimens housed in Texas herbaria provide a base of hard data that can be used for these floristic summaries and any study dealing with Texas plants.

Our Herbarium Specimen Browser uses, as a source database, the results of an initial data-gathering pass over the Tracy Herbarium's collection (which is still in progress at this time). At present only specimens collected within Texas are being recorded. For each of those, the following items are being recorded: accession number and source herbarium, collector's name, a collector-specific number for the specimen, date of collection, county of collection, and scientific name (along with some special codes relating that name to a global taxonomy). Future data-gathering passes will involve specimens not from Texas, data in annotations, and images of the plants themselves.

3. Implementation and functionality of the Specimen Browser

A hallmark of our working group's Web tool development has been the use of the public-domain information retrieval system MG [Witten 94]. MG's collection construction programs take sets of arbitrary ASCII "documents," compress them, and produce indices to facilitate querying. MG's querying programs then allow Boolean, ranked, and specific document (i.e. by document number) queries, returning results in a variety of forms, ranging from fully decompressed documents to lists of document numbers.

Our tools could be said to make use of MG's full-text retrieval facilities to emulate the query functions of a relational database. "Documents" are formed from a table's individual records; each field is prefixed with a unique string to form (in most cases) a "word" which the full-text retrieval system can search for. As a result, one can retrieve the "records" containing desired field values by retrieving documents containing desired "words".

For applications such as ours where database updates are infrequent, MG is much more convenient to use than a standard database system. Since the collections are read-only, much of the overhead caused by transaction management and concurrency facilities is eliminated. Also, the retrieval system is optimized heavily with regard to query speed by moving much computation into the collection construction phase.

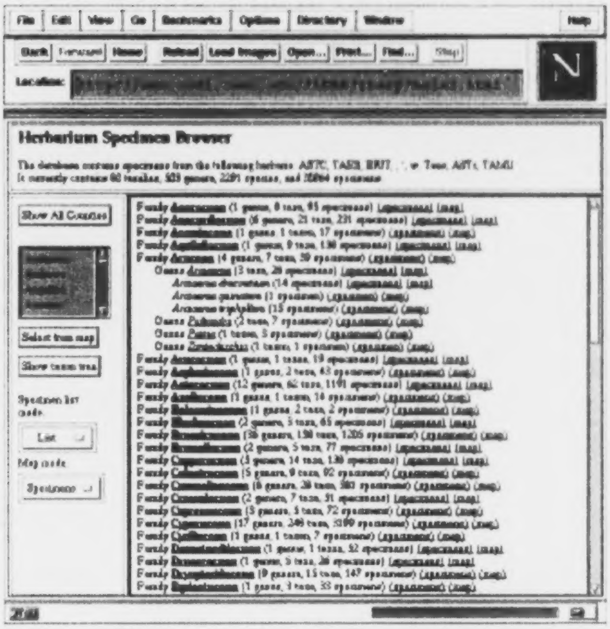


Figure 1: the Specimen Browser in use

[Fig. 1] shows the Specimen Browser in use. The top frame, which is static, contains a title and some information about the current database being viewed: the herbaria from which the specimens are drawn, the number of specimens, and the number of taxonomic families, genera, and species that those specimens are part of. The frame on the left contains a number of controls which are used to change what is displayed in the frame on the right. Initially, that frame, generated by a CGI program, consists of a view of the database at the family level, listing each family represented by specimens, and for each family, the number of genera, species, and specimens contained in it.

The family names in this display are HTML anchor sites. Selecting one of them causes an "expansion" of the display to show a listing of the genera (represented by specimens) contained in that family; one can then select one of the genera to see a list of species in the genus. Selecting an already expanded item causes its "contraction". [Fig. 1] shows the results of expanding the family Araceae, and within that family, the genus *Arisaema*. Selecting "Araceae" again would cause all sub-items under it to disappear.

The Specimen Browser provide a facility for filtering the displayed list by county of collection. Two methods for doing this are available. The first is through the use of two controls in the control bar - the "Show All Counties" button and the list box with Texas county names. Selecting (or deselecting) one or more counties in the list box causes an immediate update of the contents of the left frame. Any expanded items are still shown as expanded, but only families, genera, and species represented by specimens from the selected counties are shown; also, the totals indicating numbers of genera, species, and specimens are updated to reflect the restriction to the chosen area. (One can deselect all list items, returning to the "all counties" display, by pressing the "show all counties" button above the list.) Coordination between actions on the list and updates is done using Javascript functions.

Another method for county filtering is graphical. Pressing the "select from map" button causes a map of Texas to appear in the right frame, with the currently-selected counties colored in. Clicking a county on the map will cause the corresponding entry in the list to be selected or deselected appropriately, as well as updating the map; clicking a name in the list will cause the map to be updated in an analogous way. In this manner, one

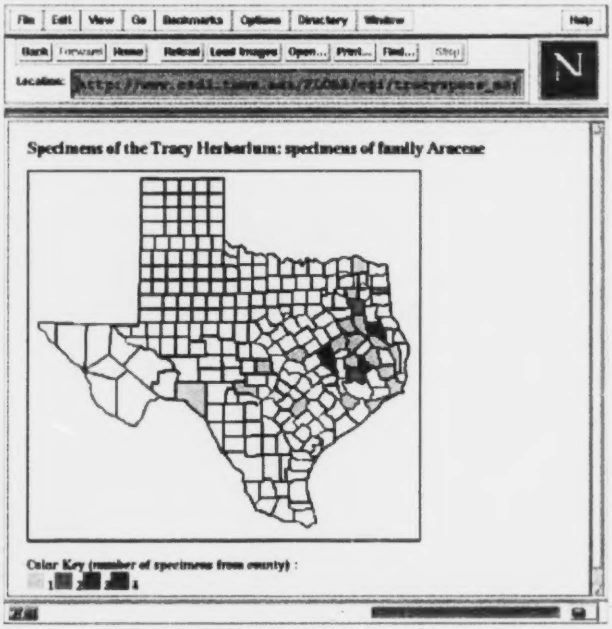
BEST COPY AVAILABLE

can build up a region of inquiry; when finished, pressing the "show taxon tree" button will redisplay the list of items, updated appropriately with respect to the new list of selected counties.

Simply displaying which families, genera, or species are located in a given set of counties would be very straightforward using a Boolean search. Generating running totals of specimens and other taxonomic categories is not so easy. It is not feasible to precompute them, since, there being 254 counties in Texas, this would require precomputing 2^{254} totals for every item. Instead, by sorting the "documents" in the MG collection in depth-first search order and precomputing lists indicating what categories cover what document ranges, the system can perform something like the SQL "select - group by" statement with the full-text retrieval system.

Each item in the list of families, genera, and species has a "specimens" link next to it. This is used to access detailed information on the specimens representing that item. The control bar contains a specimen list mode selector, which can be set to either "list" (the default) or "full data". Selecting a "specimens" link on the list frame invokes a Javascript function which examines the state of the specimen list mode selection item to determine the exact form of the URL to fetch. Requesting a specimen list in "list" mode yields a bulleted list of specimens, listing, for each specimen, its source herbarium and accession number, scientific name, collector, and county of collection. Each item in the list is a link to a "full data" display of all information about that specimen. If one requests a list in "full data" mode, full data for all specimens in the chosen group are shown, bypassing the intermediate list.

Each item in the list (on the main display) also contains a "map" link. Again, it uses a Javascript intermediate function to determine whether to show a map of the density of specimens throughout Texas or a density of species (they may differ, as more than one specimen for a species may exist for a county). [Fig. 2] shows the map of specimens of the family Araceae. The individual colored counties in the map may be clicked to display a list of the specimens from that county using the formats just described.



Specimens of the Tracy Herbarium: specimens of family Araceae

Color Key (number of specimens from county) :

Figure 2: Mapping specimen density

Most of the Web tools the working group has developed have included a clickable map feature. The maps are generated from a file representing the connected regions of the map in a run-length encoding (i.e. a list of *(region, number of pixels)* pairs representing the map as a left-right top-down raster scan) which are also used

BEST COPY AVAILABLE

to easily map *(x, y)* coordinates to regions without bounding polygons and winding rules. We believe this technique has a great deal of applicability to "irregular" image maps of all kinds, and appears to be much faster than using a full-fledged GIS system to generate the maps.

To achieve greater efficiency in the construction of the maps, another MG collection is generated from the specimen database, this time with the records sorted by county. Document numbers are retrieved via the query mechanism in the same way as described above for the main list, but the groups formed are county clusters rather than taxonomic categories. Certain specimen records are specially tagged as representatives of their species to make species-density mapping easier, by insuring only one "representative" exists per county.

4. Philosophical points behind the Specimen Browser

The following are some general points of philosophy we feel this tool exemplifies and which other designers may find useful.

Overviews and filtering. Much of our past and current work in mapping geographic distributions is motivated by the desire to give biologists meaningful overviews of large quantities of data. In this sense our work has an affinity with other digital library projects such as the Visible Human project [North 96]. The idea is to provide a general overview with allows the discernment of global patterns, coupled with the ability to quickly investigate details if desired.

The idea of the expanding, contracting, and filterable list (similar to Nelson's notion of *stretchtext* [Nelson 93]) came about as an attempt to realize this. The initial family-level overview allows one to see how specimens are distributed through the collection by family. Interesting families can then be expanded if desired and the resultant subtotals displayed. If one wishes to restrict one's attention to a specific geographic area, one can do so while still maintaining the context of one's attention to particular taxonomic items.

Viewers looking for generalities should not be forced to rely on their own memories. This motivated the implementation of the "list" versus "full data" options for viewing specimen sets. The list option allows one to look for certain general patterns, such as preponderances of collectors, without having to page through large amounts of other data. The full data option, however, allows one to see everything that is recorded about small sets, rather than forcing one to visit each specimen in turn via the list and remember the details.

It is not only important to make patterns visible, but also to avoid the impression of false patterns. Initial experiments with our maps used red and green for the high and low ends of ranges, with a blending to indicate the middle. Unfortunately, this created a midpoint color which had greater visual salience than either endpoint, creating false impressions. The effect disappeared when we switched to a single-color scheme. (Bertin's work [Bertin 81] [Bertin 83] contains many useful guidelines for map designers regarding what can and cannot be signified by color, value, shape, etc., and how those variables should relate to the actual data to avoid false patterns. Similar insights can be gleaned from the work of Tufte [Tufte 83].)

Regularity. The displays in our system are very rich in links. This gives the impression of an extensive information field which viewers can explore in an unrestrained manner. However, we avoid disorientation by having links lead to destinations (or trigger actions) in a uniform matter. In addition, link sources are uniform as well - simple rules indicate if a link should be present and are never violated. (They are thus instantiations of what DeRose calls *annotation*, as opposed to *associative*, links [DeRose 89].) This is not to say that all designers should attempt to impose uniformity on their information spaces, but it demonstrates that Web systems are suitable for constructing tools to explore detailed information spaces with regular contours.

Javascript as a simplifying mechanism. The Specimen Browser blends together static items, CGI calls, and Javascript in nontrivial ways. However, the overall effect is one of simplification. Consider that the four options of specimen list, full specimen data, species density map, and specimen density map are implemented using two controls on the control bar and two links per item. Without Javascript this would require four links per item, unnecessarily increasing screen clutter. Also, Javascript allows user selections on the county list in the control bar to trigger immediate updates of the list; without it, some additional, superfluous user action would be required to trigger this.

Javascript is often used today to create flashy substitutes for standard lists of links or scrolling marquees in browser status bars. However, with careful use it can expand on HTML's limited link model and create effects using small sets of composable components that, before, would require large, cluttered lists of links.

5. Future Work

Currently the Herbarium Specimen Browser is not enabling pattern-finding as much as it could. An obvious extension would be to allow the various displayed lists to be sorted in ways other than alphabetical, for example, sorting families by number of specimens, or specimens by collector. An important point is that a variety of sorting methods should exist to allow both pattern finding as well as random access (e.g., finding a specimen by accession number). Similarly, it should be possible to filter the lists on criteria other than "county of collection", such as collector or herbarium. Maps should be able to be drawn with similar constraints.

One aspect of the information "space" that is amenable to processing but is currently not utilized is the time dimension. Time-series display of, say, the activities of a given collector represented in an herbarium might be interesting, but it is not clear how to do this in a straightforward yet effective way.

It will also be interesting to see what other kinds of searches we can perform using MG. Searching for records on multi-word fields (like collector name) is easily done. However, complex searches on date ranges, for example (such as finding all specimens collected between a pair of dates), are difficult to perform efficiently using our current date representations; in the future we will be investigating alternate representations more suited to the searches a full-text retrieval system can perform.

References

[Bertin 81] Bertin, J. (1981). *Graphics and Graphic Information Processing.* Berlin: Walter de Gruyter.

[Bertin 83] Bertin, J. (1983). *Semiology of Graphics.* Madison, WI: University of Wisconsin Press.

[DeRose 89] DeRose, S. J. (1989). Expanding the Notion of Links. *Second ACM Conference on Hypertext (Hypertext '89),* Association for Computing Machinery, Pittsburgh, PA. 249-257.

[Nelson 93] Nelson, T. H. (1993). *Literary Machines.* Sausalito, CA: Mindful Press.

[North 96] North, C., Shneiderman, B., & Pleasant, C. (1996). User Controlled Overviews of an Image Library: A Case Study of the Visible Human. *First ACM International Conference on Digital Libraries (DL '96),* Association for Computing Machinery, Bethesda, MD. 74-82.

[Tufte 83] Tufte, E. R. (1983). *The Visual Display of Quantitative Information.* Cheshire, CT: Graphics Press.

[Witten 94] Witten, I. H., Moffat, A., & Bell, T. C. (1994). *Managing Gigabytes: Compressing and Indexing Documents and Images.* New York, NY: Van Nostrand Reinhold.

Acknowledgments

This research was supported in part by the Texas Advanced Research Program (Grant Number 999903-230).

The Internet in School: The Shaping of Use by Organizational, Structural, and Cultural Factors

Janet Ward Schofield
Learning Research and Development Center
University of Pittsburgh
Pittsburgh, PA
U.S.A.
email: schof@vms.cis.pitt.edu

Ann Locke Davidson
Learning Research and Development Center
University of Pittsburgh
Pittsburgh, PA
U.S.A.
email: adavidso@cats1.admin.pps.pgh.pa.us

Abstract: This paper explores the factors that delayed, shaped, and constrained Internet use in a large urban school district. Although a substantial amount of use occurred, problems in interfacing with the district's pre-existing physical infrastructure, its bureaucratic procedures, and the culture of its schools all influenced use markedly. Infrastructure problems included difficulties retrofitting old buildings, including asbestos in school walls, and lack of needed power outlets, space, and furniture. Bureaucratic problems included incompatibility between the rigid bell schedule and the unpredictability of access to Internet sites. Finally, cultural factors including the teachers' role as dispenser of knowledge, the image of a well-run classroom as one in which students sit quietly in their seats, the tendency to emphasize basic skills and to conceptualize learning along disciplinary lines, and concerns about ensuring that the materials students access in school are consistent with community beliefs and standards also shaped and limited Internet use.

In recent years calls to connect schools to the Internet have been legion. As just one example, President Clinton has made access to the Internet for all 12-year-olds one of the standard goals he has set for schools in the U.S. Those advocating Internet access point out a wide range of possible benefits [Hunter 1992]. Yet previous research demonstrates that the mere fact that teachers have access to computers does not mean that they will use them [Cuban 1986; Schofield 1994]. In addition, the computer use that does occur is often very much influenced by the existing school environment [Cohen 1987; Schofield 1995]. Building on such insights, this paper focuses on the ways in which existing school culture and structure can delay, constrain, and shape Internet use.

The paper is based on a four year study of a National Science Foundation funded project called Common Knowledge: Pittsburgh (CK:P) -- one of four endeavors in the United States designed to serve as national "testbeds" for the exploration of the Internet's potential for improving education. CK:P's goal, at the most general level, has been to bring Internet access to teachers in the Pittsburgh Public Schools for their use as a professional development resource, and, even more importantly, for instructional purposes. Before turning to the discussion of the results of this research, we will briefly describe both CK:P and the methodology used in gathering and analyzing the data upon which this paper is based.

Common Knowledge: Pittsburgh

CK:P is a collaboration between the Pittsburgh Public Schools, Pittsburgh Supercomputing Center, and the

University of Pittsburgh. Over the past four years, CK:P has provided teachers and students in more than 60 schools with Internet access. The project has been based on the idea that teachers are the ones most suited to discover and develop the curricular uses that fit their students' needs. Thus, teachers throughout the district have been encouraged to join together with others at their schools into groups to develop proposals to submit to annual competitions which CK:P ran to select the classrooms for which it would provide Internet access. Many of the individuals participating in these groups, particularly in the first years of the project, had little, if any, experience with computers in general, or the Internet in particular. Thus, the CK:P staff provided a great deal of training and support regarding both technical and curriculum issues.

Methodology

The major data-gathering methods relevant to the issues discussed in this paper were qualitative observations, semi-structured interviews, and the collection of archival material. Since the project began in 1993, we have conducted extended and repeated observations in a wide variety of settings. This includes over 160 hours of observations in over 40 classrooms in which the Internet was being used. It also includes observations of over 125 meetings between different groups of teachers who have been involved with the project, and dozens of meetings of CK:P's educational and technical support staff. Trained observers used the "full field note" method of data collection [Olsen 1976] which involves taking extensive hand-written notes during the events being observed. All notes were made as factual and as concretely descriptive as possible.

Because interviews are so useful in providing participants' perspectives on events, over 350 semi-structured open-ended interviews were conducted with a very wide variety of individuals. This included over 100 teachers, 30 school district personnel, and 14 CK:P staff who supplied a great deal of data pertinent to the issues discussed here. All field notes and interviews were audiotaped, transcribed, and then coded using established qualitative methods [Strauss & Corbin 1990; Miles & Huberman 1994].

Archival materials, especially e-mail, were another important source of information used in this research. With the participants' permission, the research team's address was added to virtually all group mailing lists connected with the project. This allowed us to monitor most normal e-mail correspondence between members of the various groups working on this project.

Other more quantitative data were also collected when they appeared to be particularly useful. So, for example, certain kinds of usage statistics were collected from school-based file servers and surveys of teachers were conducted.

Results and Conclusions

There is no doubt that a substantial amount of Internet use occurred in the schools involved in the CK:P project. By the end of the project's fourth year, over 4,500 teachers and students had Internet accounts through CK:P. The kinds of activities that individuals engaged in were extraordinarily varied. A sense of the range and kind of usage was captured one day roughly three years into the project when participants from around the district were asked take a few minutes to let others know what they had used the Internet for that day. A collection of contributions from over 20 locations around the district created a snapshot of the kinds of CK:P activities occurring. Although it is likely that Internet activity was unusually high on this day, the kinds of activities in which people engaged seemed quite representative of the range of activities routinely observed in the schools.

High school students in French and German classes searched for information on Paris, Quebec, and Vienna using World Wide Web sites located in those countries. Students in Spanish classes communicated with individuals in Chili over Internet Relay Chat. Students from a variety of classes reported accessing sites containing career and scholarship related information. Middle school students reported having engaged in activities such as writing to pen pals in Brazil, gathering information for reports on topics ranging from sports, to World War II, to eating disorders, and posting their own poetry for feedback on this day or earlier in the year. Elementary school children engaged in activities including work on logic projects obtained from an Internet site, visiting a virtual classroom in which they read stories and posted responses, checking weather forecasts, looking at interactive on-line maps of the city to find their own street corners, and corresponding with other elementary school classrooms to get information about two artists they were studying.

However, there was also no doubt that the level of Internet usage was constrained and that the nature of Internet usage was shaped in ways that were not always consistent with visionarys' images of the Internet's functioning in classrooms or with the participants' initial plans and hopes for it. We now turn to discussing how and why this happened, starting with a brief mention of the delays and constraints that arose from working within the physical infrastructure of a large urban school district. However, the primary focus of our paper is on the organizational and social factors that delayed, shaped, and limited Internet use.

Interfacing with the Existing Infrastructure

It became evident during the course of CK:P that providing schools with workable access to the Internet was often more difficult than anticipated. As has become apparent in "Netday" activities around the country, asbestos in walls can pose a major problem. At some CK:P sites asbestos caused substantial delays. At such schools the wiring was postponed for months in order not to expose students and teachers to it. Furthermore, the layout of some buildings made it prohibitively expensive to provide high speed connectivity in the desired places. Thus, in some cases, initial plans to put certain schools on-line so that teachers and students there could readily interact with each other around a shared curricular focus were changed in ways that reflected financial and infrastructure considerations rather than educational ones. The fact that decisions about the physical location of the drops necessary to connect computers to the Internet had to be made before teachers had much experience with using computers in their classrooms also created problems and inefficiencies. Finally, pre-existing electrical outlets, telephone lines, space, and even furniture were frequently not adequate for optional use of the new computers that project schools hoped to connect to the Internet. But the fact that money was tight meant that more often than not project teachers had to work within the constraints imposed by such factors which limited Internet use.

Interfacing with the Bureaucratic Structure Beyond the Classroom

A whole range of issues that delayed and inhibited Internet use were connected to the fact that teams of teachers working with CK:P were embedded in a larger structure with its own rules and operating procedures. Some problems of this sort were exacerbated by the fact that CK:P was a grassroots project funded from outside of the district, rather than being part of the district's own set of programs. However, many would most likely have created problems and delays in any event. Problems arising in interfacing with the district bureaucracy were extremely varied in nature. To give just one example, longstanding purchasing procedures required that purchases be made from the lowest bidder meeting the specifications laid out by the district. Thus, in one instance, computers were purchased from the lowest bidder even though the machines offered for a slightly higher price by another vendor had much greater potential for subsequent inexpensive upgrades that were likely to significantly extend to the machines' useful life. Although this did not cause an immediate problem, given the rapidity with which hardware changes and the ever increasing demands for memory, it seemed likely to curtail use in the long run.

One factor that appeared to play a major role in inhibiting Internet use was the rather rigid bell schedule which shaped teachers' and students' days. The fact that students were to study a particular topic at a particular time, at least in middle and high school, meant that they could not switch flexibly to other subjects if an Internet site they were trying to access for work in one subject was too busy to allow them access. Although students could, of course, try again the next day, the possibility that on any given day access would be either impossible or impractically slow meant that teachers needed to prepare alternative plans in case Internet activities did not proceed as intended, something which was potentially quite time consuming and thus was unappealing to them.

Internet use was also greatly effected by the attitudes and behaviors of the principals at the school level. In some cases principals were very proactive in trying to create conditions conducive to productive use, providing time and other resources to the CK:P teams. In many other cases, however, competing priorities meant that decisions made at the building level undermined Internet use. For example, at one site a project selected for study as an "exemplary" use of the Internet came to a near complete halt when a new principal arrived and assigned one of the prime movers responsible for this project to hall duty during a period she had previously used for Internet activities. At another site, the principal required the adjusting of a school home page created by teachers and students so that no one outside of the school could access it, because she was concerned about the damage that could be done if materials of which she did not approve were placed there for all the world to see.

Interfacing with the Structure and Culture of the Classroom

Teachers not only function inside of a physical and bureaucratic environment, but they are also part of an ongoing culture [Lortie 1975; Sarason 1971]. A number of aspects of traditional classroom structure and culture also appeared to inhibit students' classroom use of the Internet.

The teachers' role as a dispenser of knowledge, upon which much of the basis for the teachers' authority rests, is one important aspect of traditional classroom culture. It was not infrequent in middle school or high school for teachers to discover that at least one or two of their students knew more than they did about the use of the computers and the Internet. Some teachers adjusted to this quite readily and, in fact, found ways to take advantage of it. However, many were made anxious by the situation. Not infrequently, this resulted in decreased use on their part.

Closely connected to the image of teacher as a knowledge dispenser is the traditional image of the well-run classroom as one in which students sit quietly in their seats and listen attentively to the teacher who speaks to them as a group. Because resources were limited, teachers proposing projects to CK:P knew that they could only ask for a few computers per participating classroom. The small number of computers per class meant that many teachers had to find ways to adjust their approach to instruction, unless the computers were to sit idle the vast majority of the time. Many found this transition rather difficult, not only because they had to find ways to make sure that students using the computers at any given moment did not miss material they were later expected to know, but also because use of the computers tended to lead to more movement and noise in the classroom as students went from their seats to the machines and helped each other when confronted with technical problems. Such problems limited Internet use most noticeably in classrooms in which the teachers had little interest in or experience with small group approaches such as learning stations or cooperative learning groups.

Traditional images of what counts as learning that emphasize basic skills and conceptualize students' knowledge along disciplinary lines also shaped Internet use significantly. Traditional curricular materials such as textbooks are organized by discipline and present information in concentrated and highly organized ways designed specifically for students at given grade levels. This is generally not the case with materials found on the Internet. Concerned about efficiency and about ensuring that students did not miss out on important material they would later be expected to know, teachers sometimes treated Internet activities as optional enrichment projects to be used to fill up empty time slots or to be reserved exclusively for those who had already mastered the traditional curriculum.

Much has been written about the isolation of teachers and the importance of reducing it [Lortie 1975]. Although many teachers involved in CK:P actively reached out to others beyond their schools for professional discussions, strong norms relating to the privacy of a teacher's classroom still persisted and undercut Internet use. Specifically, if a class was not using the Internet during a period, as was frequently the case even in high use environments, teachers from another room almost never asked to have access to the machine -- even when there were interested teachers in the school who could have worked quietly by themselves and not disrupted the class in any obvious way. Since there were very limited numbers of computers with Internet access in many CK:P sites, and a great many of them were in individual classrooms, this situation undercut use substantially.

Finally, teachers are well aware of the potential disruption and controversy that can arise if students are presented in school with material that their parents find objectional. Most are used to working in an environment which includes often elaborate procedures to approve textbooks and other curriculum materials. Internet use poses a problem in this regard since it is possible for students to access materials that would never pass such procedures or to strike up acquaintances with individuals who may wish to exploit them in some way. CK:P, like most Internet projects, had both parents and students sign an "acceptable use" policy which indicated that a wide variety of materials were available and delineated the kinds of uses that students could legitimately make of the Internet. In spite of this, classroom use of the Internet was greatly reduced in many instances because of teachers' concerns about the potential for students violating this policy, which led them to allow use only when an adult could directly view the computer monitor.

In summary, although much constructive use was made of the Internet in schools participating in CK:P, such use was substantially delayed, limited, and shaped in unanticipated ways by problems created by organizational, structural, and cultural factors. To achieve the full potential of Internet use in the schools, these factors will have to be addressed at the district, school, and classroom level.

References

[Cohen 1987] Cohen, D. K. (1987). Educational technology, policy and practices. *Educational Evaluation and Policy Analysis, 9*, 153-170.

[Cuban 1986] Cuban, L. (1986). *Teachers and machines: The classroom use of technology since 1920.* New York: Teachers College Press.

[Hunter 1992] Hunter, B. (1992). Linking for learning: Computer-and-communications network support for nationwide innovation in education. *Journal of Science Education and Technology, 1*, 23-34.

[Lortie 1975] Lortie, D. C. (1975). *School-teacher: A sociological study.* Chicago: The University of Chicago Press.

[Miles and Huberman 1984] Miles, M. B., & Huberman, A. M. (1984). *Qualitative data analysis: A source book of new methods.* Beverly Hills, CA: Sage.

[Olson 1976] Olson, S. (1976). *Ideas and data: Process and practice of social research* . Homewood, IL: The Dorsey Press.

[Sarason 1971] Sarason, S. B. (1971). *The culture of school and the problem of change* . Boston: Allyn & Bacon.

[Schofield 1995] Schofield, J. W. (1995). *Computers, classroom culture, and change* . Cambridge, MA: Cambridge University Press.

[Schofield 1994] Schofield, J. W. (1994). Barriers to computer usage in secondary school teaching. In C. W. Huff & T. Finholt (Eds.), *Social issues in computing: Putting computing in its place* (pp. 547-580). New York: McGraw-Hill.

[Strauss and Corbin 1990] Strauss, A., & Corbin, J. (1990). *Basics of qualitative research* . Newbury Park, CA: Sage.

Acknowledgments

The research reported here was funded by Contract No. RED-9253452 with the National Science Foundation. All opinions expressed herein are solely those of the authors and no endorsement of the conclusions by the National Science Foundation is implied or intended.

Legal Issues in Cyberspace

Alfred J. Sciarrino and Jack S. Cook
John Wiley Jones School of Business
State University of New York at Geneseo, USA
Sciarrin@uno.cc.geneseo.edu, Cook@uno.cc.geneseo.edu

Abstract: Today many areas of law are being shaped and changed not in terms of decades and years as in the past, but in months and days. Information concerning such changes can now be accessed in minutes. This rapid legal transformation is not merely caused by the escalated use of computers, but more importantly by the development of a new concept of time and space, called cyberspace. This paper discusses cyberspace, the major decision by the Supreme Court (which greatly expands the power and influence of the computer), and attempts to classify the major legal areas in the United States that are, or will be most affected in the future.

Introduction

Today, all that is necessary to access the Internet is a computer and a modem. The easiest way to connect is through one of the major national online services, such as America Online, or through local telephone companies. These services are not free, but the charges are currently relatively inexpensive. Once online, through the use of browsers such as Netscape Navigator and Internet Explorer, and search engines such as Infoseek, Yahoo, and WebCrawler, searching for information on the World Wide Web (WWW) is quick and easy.

In addition, once connected a person may also "chat" on almost any subject in real time with other people through chat rooms, or leave messages through bulletin boards, or by email. These are just a few of the many ways to communicate with others on the Internet.

With millions of people independently accessing the Internet, a major question is can it be monitored? The answer is that presently it cannot. As the lower federal court stated in ACLU v. Reno, "[n]o single entity -- academic, corporate, governmental, or non-profit -- administers the Internet." And, constitutional questions aside, there is currently no technology available which would allow the Internet to be centralized or controlled by a single entity. Whether such technology ever will be available in the future is debatable. If so, whether the Internet should then be monitored by individuals and/or entities (like the government) will become a terribly complex legal mess. Of course, the recent decision by the U.S. Supreme Court regarding the Communications Decency Act of 1996, was the opening salvo in addressing this problem.

The Communications Decency Act of 1996: ACLU v. Reno

In ACLU v. Reno, decided on June 26, 1997, the U.S. Supreme Court ruled in a landmark decision that the Communications Decency Act of 1996, passed by Congress to police pornography on the Internet, was unconstitutional. By doing so, it affirmed lower federal court decisions declaring the Act unconstitutional and enjoining its enforcement. Left unresolved is the issue as to whether existing or future technology can actually prevent minors from accessing pornographic sites, the cultural dilemma which Congress sought to address by legislation.

Interestingly enough, this decision was made by nine judges over fifty years of age, who by their own admission know little or nothing about the Internet or computer technology. However, in order to render the critical decision regarding the future of the Internet, the Justices recognized that they first had to be able to understand key essential components of what is called, among other terms, cyberspace. Luckily for the Supreme Court Justices, lower federal court judges had faced the very same problem earlier, and carefully distilled expert testimony regarding the technology in several written opinions. These lower court decisions made the term cyberspace less abstract and more intelligible not only for the members of the Supreme Court, but for the general public, as well.

Brief History

President Clinton signed the Communications Decency Act of 1996 (referred to here as CDA or Act) into law on February 8, 1996. On the same day, the American Civil Liberties Union (ACLU) challenged the constitutionality of the Act, and moved in Federal District Court for a temporary restraining order enjoining its enforcement. The Attorney General of the United States, Janet Reno, was made the party defendant, and by stipulation agreed not to initiate any investigations or prosecutions under the Act until a three-judge panel heard arguments. Not long after, the American Library Association, Inc. (ALA) also filed a similar action. On June 11, 1996, a three-judge panel issued a preliminary injunction enjoining governmental enforcement of the Act.

What is the purpose of the Act? In ACLU v. Reno, the government asserted that the main purpose is to shield minors from Internet pornography. The government argued that it has an interest in protecting the physical and psychological well being of minors. During oral argument before the Supreme Court, the government cautioned that the Internet is "a revolutionary means for displaying sexually explicit, patently offensive material to children in the privacy of their own homes." In fact, the government continued by saying that with a click of the mouse it "threatens to give every child a free pass to every adult bookstore and video store."

This argument is certainly compelling, and there is probably no rational person who does not want to protect minors from pornographers. The problem is, however, how can minors be protected from viewing pornography on the Internet, while still allowing adult users access? Compounding the problem is what to do with providers and web sites that are not in themselves purveying pornography, but involuntarily assisting other users in reaching pornographic sites. Also, what of web sites that are devoted to issues that are perhaps only partially pornographic. More importantly, perhaps, what of the chilling effect on Internet speech itself, speech as the ACLU has described as democratizing and speech enhancing from a distinctive forum offering worldwide conversation at little cost?

The Act was broadly tailored to provide criminal penalties for telecommunications transmissions, including those by computers, that are "indecent," or "patently offensive." However, sexually explicit speech is currently constitutionally protected under complex legal principles, when engaged in by adults. Basically, in a nutshell, pornography itself is not forbidden, only what is judged by "relevant community standards" to be indecent, patently offensive, and appealing to a prurient interest can be proscribed. The rub here is what community standards are affected when pornography is accessed from any spot on the WWW?

As a result of such a chilling effect, the ACLU labeled the Act "patently a government-imposed content-based restriction on speech," which must be struck down under current first amendment law. Furthermore, aside from definitional problems with the words "indecent," "patently offensive," and "prurient interest," and which relevant community comes into play in such a case, (not to mention how pornography is accessed to begin with), there is the problem of actually enforcing the Act. In fact, in ACLU v. Reno, the government was totally unable to demonstrate a feasible way to do so.

The court pointed out that the government was unable to show a technologically reliable way to screen the age of users. Further, the government was unable to show a technologically reliable way to segregate users by age, or by who browses in chat rooms, newsgroups, or other web forums that might contain indecent material. Finally, the court concluded that:

> [e]ven if it were technologically feasible to block minors' access to newsgroups and similar fora, there is no method by which the creators of newsgroups which contain discussions of art, politics or any other subject that could potentially elicit "indecent" contributions could limit the blocking of access by minors to such "indecent" material

> and still allow them access to the remaining content, even if the overwhelming majority of that content was not indecent.

In arguments before the Supreme Court both parties attempted to raise and then demolish the problem of age verification on the Internet. The government, while admitting the high costs for commercial users, and the "prohibitively expensive" costs for non-commercial users, for WWW verification systems that could screen for age, nevertheless argued that there were alternatives. One alternative being identification cards at a nominal yearly cost to users. However, the ACLU responded that while screening techniques might be feasible for web sites, they would not work for online chat rooms and newsgroups. News reports of the oral arguments before the Court also suggested that several of the Justices queried the government about the risk that the law could make criminals out of parents who allow children access to the Internet at home, or even of teenagers who might use the medium to discuss their sexual concerns or experiences, whether real or imagined. As the law now stands outside of this cyberspace issue, certainly adults engaging in sexually explicit conversations in public, which children happen to overhear cannot be prosecuted. Obviously, the decision in this case was a major law decision, with major impact regarding the use and further development of the Internet.

Cyberspace: Access and Administration

In ACLU v. Reno, the District Court for the Eastern District of Pennsylvania stated that in order to understand legal questions applied to cyberspace, one must first have "a clear understanding of the exponentially growing, worldwide medium that is the Internet..." It then discussed the history and basic technology of this medium. As the court pointed out, the Internet is not physical, nor tangible, but "rather a giant network which interconnects innumerable smaller groups of linked computer networks." Some networks are closed, that is not linked to another computer or networks, and some are open. Open networks are "connected to other networks in a manner which permits each computer in any network to communicate with computers on any other network in the system. This global web of linked networks and computers is referred to as the Internet."

No one can determine the size of the Internet at any given time. It is constantly growing. At the time that the case was heard, it was estimated that there were over 9,400,000 host computers, with sixty percent of them in the United States. In addition, it was also estimated that there were approximately 40 million people worldwide accessing the Internet through personal computers. Further, the court stated that governments, public institutions, not-for-profit organizations and individuals own the computers and the computer networks. "The resulting whole is a decentralized, global medium of communications -- or "cyberspace" -- that links people, institutions, corporations, and governments around the world."

Legal Limitations of Time and Space

In an astonishing display of technology, oral arguments before the U.S. Supreme Court in Reno v. ACLU appeared online the very next day, while the text of the recent highly anticipated decision appeared the very same day it was announced in Washington! During the Oklahoma bombing trial, each day's official, edited court transcript was posted the same day on the web sites of various news organizations. Furthermore, these documents were provided to Internet users free of charge. Previously, it took weeks or months for lawyers to obtain transcripts (which generally cost a fortune), and they were not readily accessible at all to the public.

Cyberspace communications are rapid, sometimes instantaneous transmissions. They can be directed at groups or individuals and transmitted over a series of redundant, decentralized, and self-maintained links between computers. Furthermore, while such communications are not generally secure, they are adaptable. That is, they have the ability to be rerouted automatically if any individual link fails. In short, a transmission over the Internet has the ability to reach its destination through any number of routes, and generally in a matter of seconds.

In an article entitled "Cybertime, Cyberspace and Cyberlaw," M. Ethan Katsh quotes Justice Brandeis's observation that the law is "limited by time and space." Katsh adds that "[m]ore than this, the law might be said to have a 'sense of place' or be 'of a place' in that there are informational places that are central to the process and operation of law." He gives as an example law libraries, and individual objects "such as books, or even artifacts, such as contracts." He might have added, of course, law offices and courtrooms.

However, as Katsh points out, cyberspace has invaded these traditional "legal spaces." Cyberspace exists outside of the law's traditional physical spaces and objects. Cyberspace destroys or expands traditional notions of spatial terms. Katsh illustrates this point by pointing out that the legal notion of privacy "not only involves individual control over certain kinds of information but employs spatial terms, such as zones of privacy, to describe its nature." He concludes that the traditional boundaries of law, including jurisdiction and even membership in the legal profession itself, are "touched by cyberspace because if there is any one message of the new media, it is that traditional boundaries, whether they be physical, territorial or conceptual, are more porous in an age where information is digital in nature."

Katsh's concern extends beyond the notions of legal space. Under such digitalization, time, he says, must now be measured differently. He cautions, though, that cybertime is "more about time frames than time limits." Legal procedure is a creature of time limits. He does not envision that new technologies will do more than "encourage the shortening of some time periods, since they do allow informational tasks to be carried out more quickly than previously." And he adds that "[c]ybertime is not simply about speeding up information-related processes but having a different sense of the past and present, of the role of the past and the value of the past, and even a different series of concerns about the future." It is not the clock that is to be replaced, it is our relationship to time that will change. As an example he states that printed works are "dated" in the sense that they are already words in the past when published, and take a great deal of time to update. Electronic works, on the other hand, while also dated, can quickly be updated. He concludes that it is not that the library is irrelevant, but that a new powerful and competing source of legal information has emerged.

The Nature of Cyberlaw

The term cyberlaw, or cyberspace law, does not identify a particular body of newly recognized law emerging independently from the statutory or common law as we know it. Instead, at present, the term cyberlaw represents more the attempt to categorize and label traditional areas of law that are most affected by the development of cyberspace, or those areas of the law that are most influential in determining the legal boundaries of cyberspace. Therefore, the cyberlawyer or cyberlawreader cannot discard traditional notions of law and justice.

As many are suggesting, use of the Internet is creating a host of new legal questions. Previous cases involving communications dealt with now familiar systems and technology, such as telephones, radio and television, cable television, the postal system, and the publishing industry. The problem now is whether the courts are to consider the Internet as a fundamentally distinct medium, separately regulated, governed by it "with its own ethics, regulatory scheme, and citizenship requirements." Certainly, unlike the other communications mediums, the Internet now allows any individual to broadcast, disseminate information, and to collaborate and interact on a worldwide stage, regardless of geographical location.

The UCLA Online Institute for Cyberspace Law and Policy divides cyberlaw into seven topic areas: freedom of expression, intellectual property, privacy, safety, electronic commerce, equity, and jurisdiction. Most of the cyberlaw action, the Institute states, is in the area of freedom of expression and intellectual property.

Initially, it is probably dangerous to limit cyberlaw to such restricted categories. Where does criminal law fit, for instance, under freedom of expression, privacy, safety? Yet the Institute's categories can serve as a starting point, so long as the reader realizes that cyberlaw is rapidly changing, and any attempts to cast cyberlaw in stone must ultimately fail.

Freedom of Expression

Freedom of expression is guaranteed by the first amendment, and incorporates such rights as the free exercise of religion, and free speech and press. Simply put, it is your right to speak, write, or publish on any subject free from governmental control. However, this right is not absolute, and as an alarming fairly recent poll showed, most Americans do not truly believe in freedom of speech except for themselves.

Limits on freedom of expression are not hard to understand. As the Supreme Court said long ago, one cannot yell "fire" in a crowded theater because it would cause a panic. Of course, one can, but the state would be free to prosecute the individual. However, except for speech that presents "a clear and present danger" to others, other speech including (it may come as a surprise) pornography between and among adults, is generally protected. Although the state cannot limit protected speech, it can pass laws and regulations respecting the time, place, and manner of its transmission. For instance, a state can require that individuals passing out religious literature at a state fair only do so from a booth, and at stated times. While it has already been stated that pornography may be protected speech, child pornography certainly is not. The principle reason for this is that society deems that the right of children to be free from sexual abuse or exploitation overrides any first amendment question involved, and rightly so.

Privacy

Information is the currency of modern commerce. Futurist Daniel Burrus, author of "Technotrends," believes information increases in value as it is shared. On the other hand, information is of limited use if hoarded or in the possession of someone who does not understand its value. How information is shared is important. Participants in a conversation often communicate complex ideas by inflection or nuance. Oftentimes, conveying complex ideas is much more difficult with email. With email, a correspondent is limited to only words and must ascertain precisely what information is known and what is needed.

Electronic Communications Privacy Act (ECPA)

For those interested in locating a truly private means of communicating information, email should not be one of the alternatives considered. Although email offers some interesting and indisputable benefits, the law provides little protection of email privacy. With respect to email, the Electronic Communications Privacy Act (ECPA), a federal law passed in 1986, is most relevant. The purpose of the ECPA was to expand the federal eavesdropping law that was passed in 1968 to cover email and other emerging media. The ECPA classifies protected communications into three categories: oral, wire and electronic. The Act defines wire communications as any wire-borne communication involving voice. The distinction between wire and electronic is critical since the ECPA affords far greater protection to wire-borne voice communication. Did the drafters of the ECPA believe phone calls are more intimate and deserving greater protection than email? We will never know but nevertheless the legal distinction does exist.

The legal distinction between wire and electronic communications is worthy of further exploration. A federal prosecutor who wishes to wiretap a suspect's phone must first seek the approval of the U.S. attorney general before the request can even be submitted to a federal judge. Phone-taps are only granted in connection with the investigation of certain ECPA-designated federal crimes. On the other hand, a prosecutor who desires a warrant to eavesdrop on electronic communications does not need the approval of the Department of Justice. In fact, suspicion of any federal felony is all that is needed to seek such a warrant.

The ECPA also makes a distinction between seizure of "stored" communications and "live" eavesdropping, which affords more severe penalties to the latter. In general, the ECPA creates a separate and lesser set of protections for stored communications. Realistically, it is much easier to gain access to email once stored than it is to intercept a message during transmission. Hence, the flaw with this distinction is that email will almost always be classified as a stored communication, except in those very rare occasions where a message is intercepted during transmission.

The ECPA is filled with hidden ambiguities resulting from its complicated legislative history. Until recently, only the electronic communications category was applicable to online users. With the introduction of telephone-emulation packages on the Internet, wire-borne voice communication is now technologically feasible on the Internet. Furthermore, it is possible to attach an audio file to an email message. How the courts apply the ECPA to this type of voice-communication has yet to be seen.

Steve Jackson Games vs. U.S. Secret Service

Judicial interpretation of the ECPA continues but the case most cited is Steve Jackson Games vs. U.S. Secret Service. In that case, the Secret Service executed a search warrant at the offices of a game publisher. Pursuant to its investigation, the Secret Service seized a computer that served both as a development platform and a server for a public bulletin board system (BBS). Contained within the memory of this computer was the email of some 300 BBS customers, none of whom were the targets of the probe.

Before continuing the description of the circumstances surrounding this case, a brief explanation of some of the provisions of the ECPA are in order. The ECPA sets standards for government seizures and makes it a crime to obtain, alter or prevent access to stored communications without authorization. Furthermore, it creates a private right of action that lets victims of unauthorized invasions bring civil suits for money damages. As stated earlier, the ECPA makes a distinction between stored communications and communications in the process of transmission. The statutory minimum for invasions of stored communications is $1,000 whereas it is $10,000 for victims of illegal "interceptions." In this case, the distinction made a substantial difference monetarily to the plaintiffs.

Steve Jackson Games and its customers filed a civil suit claiming that the government's action violated the ECPA because the Secret Service took none of the preliminary steps required by the ECPA. The court agreed but declined to rule the email had been intercepted; rather, the court concluded the email had reached its destination and hence, it was stored communication. As a result, each plaintiff received only $1,000 -- the statutory minimum for invasion of stored communication.

Summary

As more people come on-line, invariably new legal circumstances will arise. This paper briefly addresses just a couple of issues thus far encountered in cyberlaw. It is imperative that members of the legal profession be aware of the opportunities and limitations created by the Internet. The purpose of this paper is to shed some light on legal issues in cyberspace.

References: available upon request

The Intranet as a Cognitive Architecture for Training and Education: Basic Assumptions and Development Issues

Ahmed Seffah(1) and Robert Maurice Bouchard(2)
(1) Computer Research Institute of Montreal
1801, McGill College, Montreal, Quebec, Canada H3A 2N4.
Phone: 514-840-1234 Fax : 514-840-1244 Email: aseffah@crim.ca http ://www.crim.ca/~aseffah
(2) McGill University, Department of Educational Psychology and Counseling
Applied Cognitive Science Research Group
3700 McTavish, Montreal, Quebec Canada H3A 1Y2.
Phone: 514-398-4914 Fax : 514 398 6968. Email: inbo@musicb.mcgill.ca

Abstract: In this article, we make basic assumptions regarding the development of an intranet architecture that will actively promote the cognitive apprenticeship of a new community of learners. We consider the intranet as a dynamic and virtual environment in which individuals may communicate, share resources, and reciprocally generate and organize learning strategies leading to knowledge and self efficacy. First, we describe our proposed architecture supported by an exemplar called SAGE-ISO. Secondly, we highlight several cognitive variables that can act as building blocks towards an efficient intranet foundation. We close our discussion with a brief overview of development issues regarding Internet/intranet technologies and tools.

Training/Education and Internet Technologies/Tools

Traditional computer-based instruction (CBI) has the capacity of being dynamically transformed by Internet/Intranet Technologies and Tools (ITT). Independent of computer hardware, this platform can support just-in-time media-rich content, as fresh as the moment and modified at will. It also offers a flexible structure allowing self-directed, self-paced instruction on any topic, capable of being supported by adaptive remedial and assessment strategies.

ITT is also an ideal vehicle for effective courseware delivery to individuals anywhere in the world at any time. Advances in computer network technology and improvements in bandwidth are presently introducing unlimited point-to-point as well as multi-point multimedia on-demand. Web browsers supporting 3-D virtual reality, animation, interactive transactions and conferencing are also presenting unparalleled training and education opportunities. Web-based performance systems can also actively support today's demanding workforce by integrating information systems, job aids as well as anchored instruction into unified systems available on demand.

The current focus of Web-based development is concerned with learning how to use available Internet technologies and tools as well as organize content into well-crafted teaching systems. The Web is a vehicle for the distribution of resources as well as a medium of expression-representation with its own specificity. Training designers are presently struggling with issues of user interface design and programming directed at high levels of interaction. Unfortunately, there are very few examples of good Web-based design available on the public Internet. As instructional designers and courseware developers learn to write and produce Web based resources, and as training vendors come to realize the overwhelming advantages of this delivery method, we can expect an explosion in training offerings available over the public Internet and corporate intranets.

The paper is organized as follows : Section two describes our proposed architecture supported by an exemplar called SAGE-ISO. Section three addresses several cognitive variables that can act as building blocks towards an efficient intranet foundation. Section four presents a brief overview of development issues regarding Internet/intranet technologies and tools.

The Intranet Environment Architecture

Present Accomplishments

Viewed from the end-user's perspective, [Fig.1] illustrates an architecture capable of supporting training and educational Internet/intranet transactions.

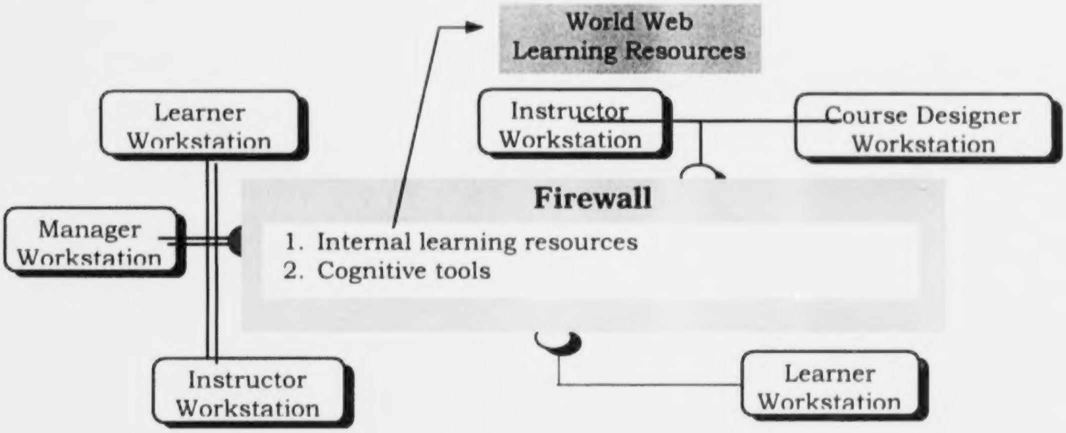

World Web
Learning Resources
Learner Workstation
Instructor Workstation
Course Designer Workstation
Firewall
Manager Workstation
1. Internal learning resources
2. Cognitive tools
Instructor Workstation
Learner Workstation

Figure 1. The Architecture of the Intranet Training Environment

SAGE-ISO is an exemplar that we have developed. It includes the following cognitive tools :

1. Browsing for information regarding ISO 9000 standards as well as a company's quality system.
2. Advising the user on deploying the quality procedures. Information supplied by the advisor tool concerns the main steps to be accomplished and the documents to be used. This tools aims also at reducing or avoiding errors due to an incorrect use of procedures.
3. Training through a set of learning resources. Each learning unit enables the user to attain a coherent and generally unique instructional goal.

The term learning resource has often been used with various meanings. Specifically, we make a distinction between two kinds of learning resource units:

1. Units promoting understanding or dispensing further information. Examples of these units include HTML documents, videos and simulations. The learner exploits these resources to achieve a greater understanding of the domain knowledge.
2. Units describing problem-based learning activities, cases studies and demonstrations. These units enable the learner to attain a coherent and generally unique instructional objective among those specified in the curriculum. Specifically, they refer to course objectives and their links with the appropriate learning resources [McCalla 92, Halff 88].

BEST COPY AVAILABLE

Future Considerations

Our next step is to design and develop cognitive tools that will actively support the Intranet learning process. Our challenge is to discover what that means in context. Computer-based instruction is traditionally rooted in well-defined course goals and objectives. They in turn are clearly stated in succinct terms associated with behavioral outcomes that are themselves directly related to corresponding sequences of instructional events. The end result is that hopefully, the user will experience a meaningful and satisfying learning outcome. Cognitive Psychology is concerned mostly with problem solving and the understanding of complex cognitive skills. In terms of learning, this is in direct contrast to memorizing large block of data or simply accomplishing procedural tasks. Learning is viewed as a constructive process where changes occur to the internal representation of knowledge [Wildman 81] Instead of learning responses to an event, the cognitive experience emphasizes learning the information [Shuel 87]. We advance the premise that the intranet as a dynamic and virtual environment in which individuals communicate, share resources, and have the potential to reciprocally generate and organize learning strategies is in need of a new, non traditional model for learning. Our understanding is based on a fundamental and yet uneasy compromise between traditional courseware delivery and user acceptation and a constructivist paradigm which is concerned with how we construct knowledge from our experiences, mental structures and beliefs that are used to interpret objects and events. The next section, examines our efforts in establishing a coherent set of tools for this new model.

Cognitive Foundations

Our basic assumptions regarding the implementation of cognitive technology within this setting is that:

(a) **learning should be an active and not a passive experience.** Inert knowledge [Whitehead 29] is the process in which students acquire facts that they cannot access and use appropriately. Passive learning is the contraposition of intentional self-directed learning. [Brown 77] characterizes these two conditions as diseases of schooling and unfortunately, they are still in evidence as ongoing learning strategies. Memorizing large amounts of information and resources currently available in SAGE-ISO, is not what learning is all about. Learning theory predicts and studies have demonstrated that immediate and frequent feedback, cooperative learning and well structured exposition information and data can improve the learning process [Briggs 95]. Our challenge is therefore to create an intranet setting that will be a dynamic learning environment that will encourage reflective practice among students and teachers [Brown 1992]. Brown characterizes classrooms as work sites that are inhabited by students who perform assigned tasks under the management of teachers into a community of learners where the same students will be given significant opportunity to take charge of their own learning.

(b) **Learning can be facilitated by situating the learner within an authentic setting**. [Brown 89] has speculated that activity and situations are integral to cognition and learning. Our intranet environment describes a cognitive technology that will empower both the student as an individual and as a working contributor with other participants in search of virtual learning outcomes. SAGE-ISO is an example in which, Metacognitive strategies for learning and remembering are encouraged within the world of ISO 9001 Standards. Metacognitive skills are the strategies one uses to learn and to solve problems. SAGE ISO accomplishes this by (1) presenting the student with contextualized resources and information, (2) extending those resources and information leading to an acquisition of knowledge, (3) communicating this augmented knowledge and understanding with other participants within the course parameters.

(c) **Learners should take charge of their own learning**. [Rumelhart 80] reports a cognitive theory of learning that states that people successfully solve problems by developing mental models of the problem domain and applying their models at hand. [Shuel 87] asserts that learning is viewed as process of building, testing and refining these models until they are reliable in new problem solving situations. [Bandura 77] social cognitive theory of human learning and functioning has also proposed the concept of reciprocal dynamic interactions. Our proposed intranet setting invites the individual to continually interact with a virtual environment that offers different ways of representing and modifying information and resources into knowledge.

Development and Technological Issues

[Fig.2] illustrates the implementation of our proposed architecture based on current commercial tools and technologies and the conventional Internet infrastructure.

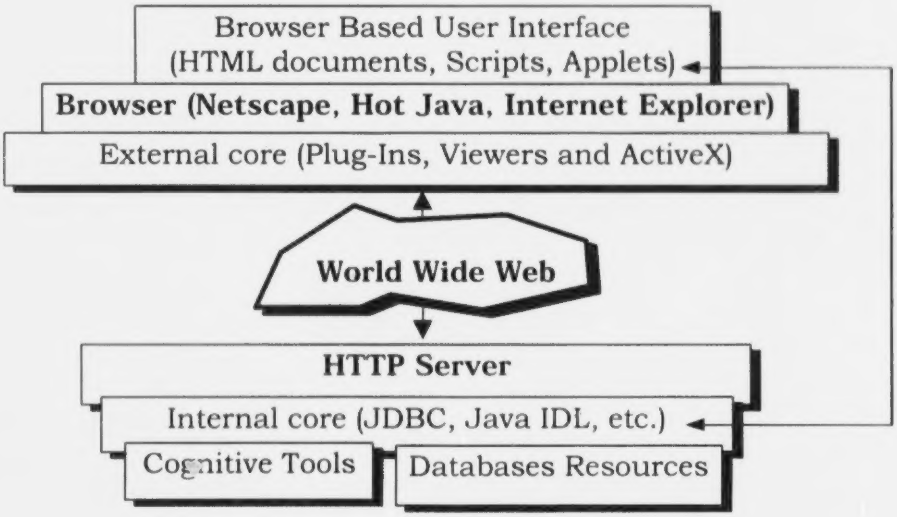

Browser Based User Interface
(HTML documents, Scripts, Applets)
Browser (Netscape, Hot Java, Internet Explorer)
External core (Plug-Ins, Viewers and ActiveX)
World Wide Web
HTTP Server
Internal core (JDBC, Java IDL, etc.)
Cognitive Tools
Databases Resources

Figure 2. Proposed Internal Architecture

This implementation includes the following components:

1. a browser-based user interface consisting of several applets. As an example, SAGE-ISO exploits the *View Engine* applet which permits information and resources to be viewed from multiple perspectives. This functionality is extremely useful in encouraging the end-user to take charge of their own learning by personalizing their understanding of the content.

2. The external core is a set of Plug-ins, Viewers and ActiveX that extend the capabilities of the browser. This technology gives the end-user the possibility of manipulating different types of resources with the same method. As an example, in SAGE-ISO clicking on a hyper link in an HTML resource, can lead the end-user to discover several other resources such as a slide show within Power Point, a multimedia tutorial or a selected assessment strategy.

3. A multi layered, object-oriented repository which stores all information and resources.

4. The internal core is a set of drivers that bridge the resources database and remote objects with the HTTP server and the user interface.

[Tab.1] summarizes the major components with their appropriate tools and technologies.

COMPONENTS	TOOLS AND TECHNOLOGIES
User Browser Interface	1) Applets 2) HTML 3) Client Script generally written in JavaScript
Web Browser	4) Internet Explorer 5) Netscape
External Core	6) Plug-ins 7) Viewers 8) ActiveX
Web Server	9) Internet Information Server 10) Netscape Enterprise Server
Internal Core	11) CGI Scripts mostly written in Perl. However it possible to use an another language such as C, C++ or Java 12) ISAPI (Internet Server Application Programming Interface) 13) JDBC such as DbAnyWhere

Table 1. Examples of Tools and Technologies

There are however, several limitations associated with the current technology and tools. They are viewed as serious obstacles to an effective and efficient intranet training and education environment. Examples of these limitations are (see also [Seffah 97] for further information):

- Applets have many restrictions and need much time for downloading.
- Servers need to be concerned re. trivial user interactions.
- Logically independent parts cannot run independently to serve multiple clients.
- The browser's BACK and FORWARD button mechanism is in direct conflict with a cognitive technology that encourages self-directed learning.
- Information and resources that are available outside the system cannot easily be cognitively integrated within the intranet training and education environment.

Conclusion

In this paper, we have presented the foundation of an intranet environment that actively promotes a Metacognitive technology approach to complex training and educational problems in need of interconnected solutions.

Some of the foundations issues discussed in this paper have been implemented and repeatedly tested through our exemplar SAGE-ISO. Many companies have also validated our cognitive architecture. Our preliminary conclusion regarding this new architecture is encouraging us to continue our research and development towards a cognitive perspective.

References

[Bandura 77] Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, N.J: Prentice Hall.
[Briggs 95] Briggs, R. O., Ramesh, V., Romano,N. C., & Latimer, J. (1995). The Examplar Project: Using group support systems to improve the learning environment. *Journal of Educational Technology Systems*, 23(3), 277-291
[Brown 89] Brown J. S., Collins A. & Duguid, P.(1989) Situated cognition and the culture of learning. *Educational Researcher*, 18 (1), 32-41

COMPONENTS	TOOLS AND TECHNOLOGIES
User Browser Interface	1) Applets 2) HTML 3) Client Script generally written in JavaScript
Web Browser	4) Internet Explorer 5) Netscape
External Core	6) Plug-ins 7) Viewers 8) ActiveX
Web Server	9) Internet Information Server 10) Netscape Enterprise Server
Internal Core	11) CGI Scripts mostly written in Perl. However it possible to use an another language such as C, C++ or Java 12) ISAPI (Internet Server Application Programming Interface) 13) JDBC such as DbAnyWhere

Table 1. Examples of Tools and Technologies

There are however, several limitations associated with the current technology and tools. They are viewed as serious obstacles to an effective and efficient intranet training and education environment. Examples of these limitations are (see also [Seffah 97] for further information):

- Applets have many restrictions and need much time for downloading.
- Servers need to be concerned re. trivial user interactions.
- Logically independent parts cannot run independently to serve multiple clients.
- The browser's BACK and FORWARD button mechanism is in direct conflict with a cognitive technology that encourages self-directed learning.
- Information and resources that are available outside the system cannot easily be cognitively integrated within the intranet training and education environment.

Conclusion

In this paper, we have presented the foundation of an intranet environment that actively promotes a Metacognitive technology approach to complex training and educational problems in need of interconnected solutions.

Some of the foundations issues discussed in this paper have been implemented and repeatedly tested through our exemplar SAGE-ISO. Many companies have also validated our cognitive architecture. Our preliminary conclusion regarding this new architecture is encouraging us to continue our research and development towards a cognitive perspective.

References

[Bandura 77] Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, N.J: Prentice Hall.

[Briggs 95] Briggs, R. O., Ramesh, V., Romano,N. C., & Latimer, J. (1995). The Examplar Project: Using group support systems to improve the learning environment. *Journal of Educational Technology Systems*, 23(3), 277-291

[Brown 89] Brown J. S., Collins A. & Duguid, P.(1989) Situated cognition and the culture of learning. *Educational Researcher*, 18 (1), 32-41

[Brown 77] Brown, A. L. (1977) Development schooling and the acquisition of knowing about knowledge. *Schooling and the Acquisition of Knowledge*. Hillsdale, N. J: Lawrence Erlbaum Associates Inc. pp241-253).

[Brown 92] Brown, A. L. (1992) Design experiments: theoretical and methodological challenges in creating complex interventions in clasroom settings. *Design of the Learning Sciences*. 2(2), 141-178

[Halff 88] Halff, H. (1988). *Curriculum and instruction in ITS*. In foundation of ITS. Hillsdale, NJ: LEA. 79-108.

[McCalla 92] McCalla, G.I. (1992). The search for adaptability, flexibility and individualization: approaches to curriculum in ITS. *Adaptive learning environments: foundations and frontiers*. Berlin: Springer Verlag, M. Jones and P.H. Winne.

[Rumelhart 80] Rumelhart, D.E. (1980). Schemata: the building blocks of cognition. In theoretical issues in reading comprehension. Erlbaum, Hillsdale, New Jersey. 33-58

[Seffah 97] Seffah, A., Khuwaja, R., Desmarais, R. (1997). Design issues of a web-based training and assessment system: from a monolithic to a distributed architecture. *International conference on computer in education (Asia chapter)*. Association for the Advancement of Computing in Education, Charlottesville, VA.

[Shuel 87] Shuel, T J. (1987). Cognitive conceptions of learning. *Review of Educational Research*. 56, 411-436

[Tate 95] Tate, R.G. 1995. At the crossroads: higher education and technology development. *Technos Winter*, 4(4)

[Wildman 81] Wildman, T.M. (1981). The cognitive conception of learning. *Review of Educational Research*. 56, 411-436.

[Whitehead 29] Whitehead, A. N. (1929). *The Aims of Eduacation*. New York: McMillan

An Architectural Framework for Developing Web-Based Interactive Applications

Ahmed Seffah(1) and Ramzan Ali Khuwaja(2)
(1) Computer Research Institute of Montreal
1801, Avenue Mc Gill College Suite 800, Montreal, Quebec, Canada H3A 2N4
Phone: (514) 840-1235 Fax : (514) 398-1234
E-mail: aseffah@crim.ca Web : http ://www.crim.ca
(2) Computing Technology Laboratory, Nortel Technology, Ottawa, Canada
Phone: (613) 765-6551 Fax: (613) 763-4222
Email: ramzan.khuwaja@nortel.ca

Abstract: The Internet and its languages JAVA and HTML offer major opportunities to develop a new generation of software applications. These new applications, are highly interactive and platform-independent, and run on a client Web browser across a network. The purpose of this paper is to discuss the software engineering issues associated with this new generation of software applications. Our main goal is to define a framework for developing such applications. The framework is based on: a reusable object-oriented architecture and a software structure which support integration at the presentation, control, data and process levels. In this paper we describe what we have learned during the development of World Wide Web Based applications during the last year.

Definition and Evolution of Web-Based Interactive Application

The first generation of World Wide Web-Based Applications (WBA) has been a succession of pages of HTML text and graphics which could be delivered virtually anywhere. The interaction was essentially a static navigation through hypertexts links. With the extensions of the HTML language, especially the introduction of the form and frame tags, the user has been given a simple mechanism to exchange information with programs and databases via the HTTP server.

In this context, Java provides an advanced framework for building a new generation of WBA. The main features of WBA developed in Java are:

- The application is automatically distributed across a network.
- User interfaces are platform-independent and support extensive interaction.
- Applications run in a client Web browser on any hardware platform and any operating system.

Web-based applications are also becoming large and complex. As a result, they require the same kind of project management, systems analysis and design, and configuration management that has been in evidence and practice in traditional software applications (see also [Yourdon 96b]. This context is compelling us to think about the software engineering issues involved in proper WBA development. In this paper we present what we have learned during the development of such applications during the last two years. However before describing these issues, let us consider as examples, two Web-based applications that we developed.

Examples of Web-Based Applications

To illustrate our proposals, we will consider the two following Web-based applications that we developed. The *Library Wizard* [Seffah 97] is a Web-based intelligent system designed to provide continuous on-the-job advising and training regarding software libraries.

In addition to tools for browsing, the systems includes tools for advising and training through a set resources. The system is developed around a unique object-oriented repository which includes all the pertinent information about a library, its services, training and advising resources (examples, problems, etc.). The system is remotely accessible across the Internet and/or corporate intranets, supports any hardware platform and runs on any operating system. A friendly web-based user interface displays advice information and training resources in accordance with the user preferences and goals.

SAGE-ISO [Seffah 98] is a user task-help system for managing the quality specified in ISO 9000 standards. The system reuses the approach and tools developed in the Library Wizard system. However, the information and resources are stored in data files under the control of the HTTP server.

Lessons Learned from SAGE-ISO and Library Wizard Development

WBA Software Architecture

In this section, we focus on the limitations of SAGE-ISO and the Library Wizard architecture. We also provide foundations for an object-oriented architecture making the system easy to understand, change, test and maintain.

Intuitive Architecture

The first version of the architecture was inspired by the common architecture for Internet-based applications [Yourdon 96a]. It comprises three major components: Browser Based-User Interface, Internet Servers, and Common Gateway Interfaces and Server API which connect the HTTP server to applications and databases [Fig.1].

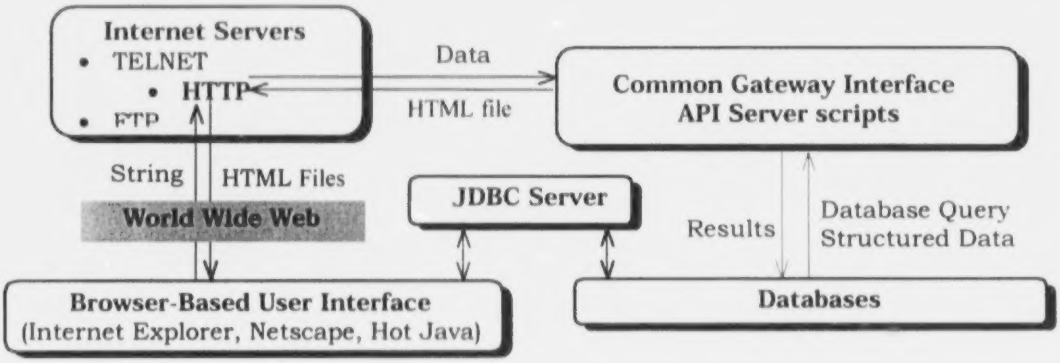


Figure 1. An Intuitive Architecture for Web Based Applications

SAGE-ISO and Library Wizard are developed around an HTTP server connected to databases. A description of SAGE-ISO and Library Wizard components follows [Tab.1].

The following three comments about the intuitive architecture must be made from the designer's point of view.

First, this architecture considers WBA as two monolithic components consisting of a Web-based user interface and server side components. It is therefore very difficult to identify which components or parts of the components can be reused to develop a new application.

The architecture does not make a distinction between the objects specific to a certain kind of Web applications and those shared by all Web based applications. Consider the following; since a communication protocol is the

BEST COPY AVAILABLE

same between all the applets (user interface components), it can be designed and implemented as a definitive design pattern [Gamma 95].

Components	SAGE-ISO	Library Wizard
HTTP Server	1. Apache	1. Netscape Communication Server
Applications Databases	2. HTML, Word, Power Point documents which are stored in file and under the HTTP server control.	2. Access database which includes information about libraries, their services and how to use them
Database Connection	3. PERL scripts	3. Java programs 4. PERL scripts
Browser Based User Interface	4. Java applets 5. HTML documents 6. Scripts	5. Java applets 6. HTML documents

Table 1. Description of the Architecture Components

Secondly, The architecture does not make a distinction between the objects specific to a certain kind of Web applications and those shared by all Web based applications. Consider the following; since a communication protocol is the same between all the applets (user interface components), it can be designed and implemented as a definitive design pattern.

Our last comment is that the architecture does not make a distinction between user interface objects and other objects. In an applet that interacts with a database, we must physically separate between code related to the look and feel of the applet and the JDBC programs. Our position is that it is important to follow the guidelines discussed in the MVC (Model-View-Controller) architecture [Goldberg 84]. This model correctly separates the interface and as a result the user interface becomes effectively portable and can be customized for each user's preferences and goals. Customization can be performed at run-time or during the design process, without modifying the other components of the WBA.

Towards an Object-Oriented Architecture

The limitations mentioned above have led us to reflect on an object-oriented architecture which could be can be reused at each stage of development, or at least uses it as a starting point for the design of some other projects.

The object-oriented architecture that we are developing is inspired from the MVC model and takes into account several design patterns.

The correspondence between MVC and the intuitive architecture is described:

Browser Based User Interface → *Views*
Servers (HTTP + JDBC) → *Controller*
CGI and API Server Scripts → *Controller*
Database → *Model*

A rigorous analysis of SAGE-ISO and the Library Wizard prototypes reports the following properties and redefines the correspondence as follows:

Browser Based User Interface → *Views + controller + model (especially when the interface includes an applet which exploit a data file)*
Servers (HTTP + JDBC) → *Controller*
CGI and API Server Scripts → *Controller + Model*
Database → *Model*

[Fig.2] presents our current vision of WBA object-oriented architecture. In this model:

- An HTTP server is a remote controller that may handle direct new incoming connections, create new client objects, decode requests, and send replies and results.
- A JDBC (Java Database Connectivity) is a connector that API defines as Java classes to representing database connections, queries, result sets, etc.
- An ORB (Object Request Broker) for Java is a connector structured to allows a Java client to transparently invoke an IDL (Interface Definition Language) object residing on a remote server. Similarly, it allows a Java server to define objects that can be transparently invoked from IDL clients.

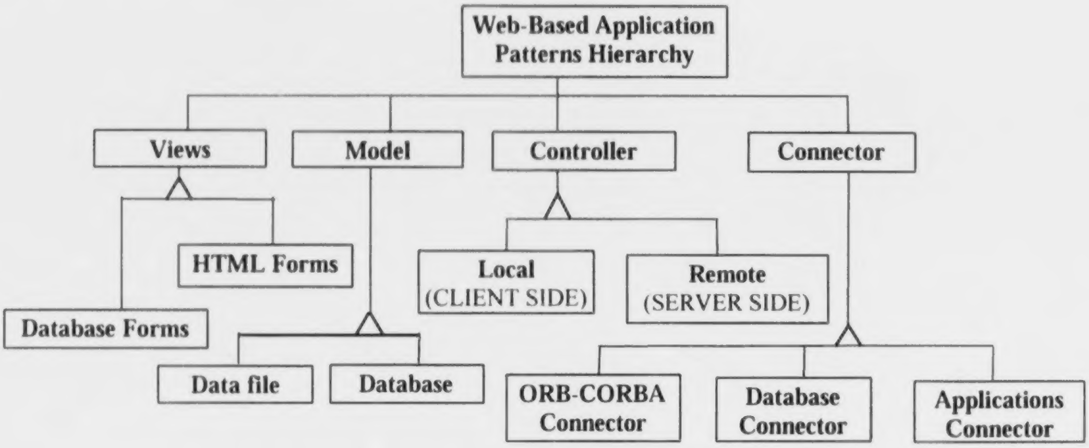


Figure 2. A simple OMT Object Model of WBA objects

The development of the SAGE-ISO and the Library Wizard prototypes shows us the importance of the following design guidelines:

- Although it is possible to invoke applets by different means, they usually must have the same access to a wide range of language capabilities. As an example, a component can access a host database, retrieve the data it needs, perform local data processing and return the results back to the host.
- User interface applets are platform independent. However well designed applets must be constructed using only common features of hardware platforms and ergonomics guidelines [Wagner 96, Jacob 96]. The resulting applets will be both cognitively and computationally efficient.
- Applets can travel from client to server and also from server to client, blurring the distinction between client and server. Applets such as our View Engine wanting to search a database spread across multiple models could dynamically send an applet to each server that will do the work. Thus, the independence between user interface and applications is guaranteed.

All these guidelines can be implemented as design patterns from which specific applets can be generated. Our object-oriented architecture must be augmented by a set of design patterns.

Web-based applications core

The core encompasses several distinct program services; these services are designed to provide a standard way to Plug-ins WBA objects at run time.

The current core

Regarding the current Internet technologies and tools, the core infrastructure is divided into two parts [Fig. 3]:

- Client side: The external core includes a set of Plug-ins that will extend the capabilities of a Web browser (e.g. to define an applet in an HTML document). SAGE-ISO uses plug-ins for all the Microsoft Office Applications (Word, Excel, Power Point).
- Server side: The internal core consists of a set of drivers that bridge existing databases, remote objects and others applications which are not necessary WBA to HTTP server and Web-based user interface. Data services defines a collection of libraries which comprise a powerful system for providing the application programmer with the ability to access and operate on data independent of its file format or its physical characteristics such as size or data type.

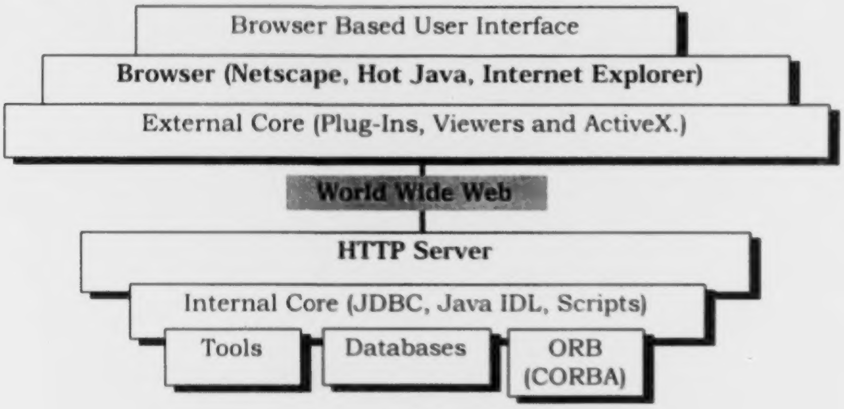

Browser Based User Interface
Browser (Netscape, Hot Java, Internet Explorer)
External Core (Plug-Ins, Viewers and ActiveX.)
World Wide Web
HTTP Server
Internal Core (JDBC, Java IDL, Scripts)
Tools
Databases
ORB (CORBA)

Figure 3. Architecture of the Web-Based Application Core

From the core to a structure for integration

We postulate that external and internal cores must support integration on the four following levels:

- Presentation integration deals with the presentation of Web Based-User Interface. Integration at this level attempts to increase productivity by allowing users' experience with previous WBA in the process to help them identify the functions of the WBA in the next stage. There are two points to presentation integration that should be noted : look and feel of the Web Based User Interface and interaction integration.
- Data integration allows a WBA to share data and data structures as appropriate with others applications that are not necessarily necessary Web-based applications. Data integration attempts to provide a single consistent source of information in which all the WBA in the work flow can use and manipulate. It can be defined with five different characteristics: inter-operability, data exchange, data consistency, non-redundancy, and data security. The development of SAGE-ISO highlighted the lack of a coherent data integration strategies within the current tools. SAGE-ISO prototypes dynamically exchange data with remote databases and project management tools. All this data are important for making strategic decisions during the quality process. The current solution implemented within SAGE-ISO, is a monolithic program which looks like «spaghetti».
- Control integration deals with how well WBA in the Internet and/or corporate intranets share functionality. Note that this type of integration may require some form of data integration as it is often difficult to share a tool's services without also sharing its data. The two sub-categories that make up control integration are provision and use. Provision integration defines how well a WBA allows other WBA to use its services in their execution, and use defines how much a WBA uses other WBA services.
- Process integration tries to measure how well one or several WBA integrate to support the entire work flow. At this level of integration, the WBA understand steps and constraints of the work flow (in SAGE-ISO the quality management process and in Library Wizard software reusability process), and they are designed to support it.

BEST COPY AVAILABLE

Our current effort is to connect the SAGE-ISO proper tools to other applications and databases in order to support the overall quality management process.

The Netscape Navigator Plug-ins API is a powerful toolbox that can be used to extend the external core. This API allows third parties to extend the Netscape Navigator with native support for new data and object types.

Conclusion

In this paper we have discussed several issues associated within the development of World Wide Web based applications. We have focused on their architecture. We have also illustrated how the evolution from an intuitive architecture to an object-oriented model can make the design and development easier and more effective. We have also examined the current Internet tools and technologies in order to implement the a WBA integration infrastructure. The current tools and technologies, which are generally based on Java foundation APIs, concentrate on the development of single components of the whole WBA and not on their composition and integration. Our proposed WBA external and internal cores are defined and designed to automatically and dynamically support composition and integration.

References

[Booch 91] Booch, G. (1991). *Object-oriented analysis and design with applications*. Redwood, Benjamin Cummings.

[Dewayne 92] Dewayne, E. and Wolf, A.L. (1992). Foundations for the study of software architectures. *Software Engineering Notes*. 17(4).

[Gamma 95[Gamma, E. (1995). Design patterns: elements of reusable object-oriented software. Addison Wesley.

[Goldberg 84] Goldberg, A (1984). *Smalltalk-80 (1984): The Interactive Programming Language*. Addison Wesley.

[Jacob 96] Jacob, K. Schmitz, P. (1996). How to build your own class libraries for the Java language. *Java One Sun's Worldwide Java Developer Conference*. San Francisco.

[Lipschutz 96] Lipschutz, R.L. (1996). Web Servers. *PC Magazine Online*.

[Seffah 97] Seffah, A., Nguyen, C.L. (1997). A web training and advising system for software libraries. *Proceeding of WebNet'97 : the World conference on WWW, Internet and Intranet*. Association for the Advancement of Computing in Education, Charlottesville, VA.

[Seffah 98] Seffah, A., André Beaudoin, A. (1998). The intranet as an environment for supporting quality assurance process: the SAGE-ISO Project. *Proceedings International Conference on the Management of Networked Enterprise*. Montreal, Canada.

[Sun 97] Java applications programming interface overview (1997). Sun Micro Systems Products. Web Site http://java.sun.com/products/apiOverview.html

[Wagner 96] Wagner, A. (1996). Designing human interfaces that make use of applets on web pages. JavaOne Sun's Worldwide Java Developer Conference(SM). San Francisco, May 29-31.

[Wong 93] Wong, W. (1993). *Plug and Play Programming : An Object-Oriented Construction Kit*. NY: M&T Books, Henry Holt Company.

[Yourdon 96a] Yourdon, E. (1996). Common architectures for Internet applications. *Corporate Internet Strategies Journal*. 2(4). Cutter Edition.

[Yourdon 96b] Yourdon, E. (1996). Internet/web-oriented application development tools. *Corporate Internet Strategies Journal*. 2(4). Cutter Eds.

Teaching Cooperative Task Using the Web

Tadié Guepfu Serge, Université de Montréal , Case postale 6128, succursale Centre-Ville,
Montréal (Québec), H3C 3J7 Canada. Email: tadie@iro.umontreal.ca

Lefebvre Bernard, Université du Québec à Montréal, Case postale 8888, succursale Centre-Ville,
Montréal (Québec), H3C 3P8, Canada. Email: lefebvre.bernard@uqam.ca

Frasson Claude, Université de Montréal , Case postale 6128, succursale Centre-Ville,
Montréal (Québec), H3C 3J7 Canada. Email: frasson@iro.umontreal.ca

Abstract: In order to build an ITS (Intelligent Tutoring System) which can teach a cooperative task, we have realized a learning system using an agent-based architecture. This system performs a « learning by doing » strategy and helps a team of learners to achieve a cooperative task. It creates agent tutors. Each of them has to assume a rôle, and has three kinds of behaviors based on three levels of reasoning : reactive, cooperative and social. A tutor uses different reasoning levels to be able to help the learner. This system is based on a modeling language of a cooperative task named MONACO_T.

Introduction

The state of the art in ITS gives some answers to these question : how knowledge is organized [Nkambou et Gauthier, 96], which kind of strategy the tutor has to use [Aimeur et Frasson, 96], how can we model the learning strategies to reuse them [Frasson et al, 96], how can we analyze student reasoning [Djamen, 95], how integrate cooperative learning and other aspects of social learning [Tadié, 96], [Chan et Baskin, 90]? But, all these goals assume that we have one student who faces the system and tries to learn an individual task or knowledge. In reality, many tasks are perform cooperatively. An example can be the task which consists on taking care of a patient in an intensive care unit. This task involves at least three roles :the doctor, the nurse and a laboratory technician.

Conceptually, what is a cooperative task ? According to [Schmidt et Bannon, 92], [Ellis, Gibbs et Rein, 91], [Terveen, 95] [Tadié, 96], a task is called cooperative if it needs two or more actors to be performed. To teach this kind of task, a learning system has to provide an environment with take into account many working-places and tools to support cooperative work. In this environment, the learning system has to coach every learner as if he was working in a classic ITS on an individual task.

In order to teach a cooperative task, we propose a system which generates a tutor for each rôle of the cooperative task. When a learner wants to learn a role, we assign him the corresponding tutor. Before beginning the learning process, all the roles must be assigned. If a role is not assigned to a learner, the tutor who has taken in charge this role simulates this one in order to allow the cooperative task to be performed.

In this paper, we will present the model we use to represent cooperative task. This model is called MONACO_T [Tadié et al, 96] and has been designed to support the learning of a cooperative task. Then, we will present the architecture of our learning system which implements the concept of tutor agents.

MONACO_T : A Model to Represent Cooperative Tasks

Because a cooperative task is a task which needs coordination between several expertises, a model of cooperative task used in a learning system should help people to know how to coordinate their actions and how to execute them. This model has also to give to the learning system tools to build a simulation of each role of the cooperative task. MONACO_T which is a task-oriented model focuses on the problem of explaining different reasoning failures and simulations of the cooperative task.

MONACO_T is organized in two levels : the static and the dynamic levels. The first one represents the decomposition of task into sub-tasks, the second one represents the rules which guides the execution of the task.

Static and Dynamic Representations

The static representation gives us a way to analyze the whole task and split it into sub-tasks. The link between a task and it's sub-tasks is the composition link. This decomposition looks like the GOMS (Goals, Operators, Methods and Selection Rules) decomposition [Card, Moran & Newell, 83]. The differences between these two models are the following :

- In MONACO_T a task is not an abstract goal, but represents a concrete action. When this task have some sub-tasks, it's action consists in the composition of the results of his sub-tasks.
- We take into account a cooperative execution of a task by allowing each task (action) to be executed by an agent.
- The static representation of a task is a tree whose root is the task, the nodes are the sub-tasks, and the leaves the elementary tasks.

The dynamic level defines the behavior of the task when a team of agents achieves it. While the GOMS model uses selection rules to determine the behavior of a task, MONACO_T builds his dynamic on the basis of the static task tree by creating in each node, three rule bases : the activation, the realization and the termination rule bases. These rules behave like pre-conditions, invariant and post-conditions rules in the domain of program proof. They can be used by a computerized tutor to determine at what time a task has to be activated, executed and terminated. They have also to synchronize the partner's actions in the cooperative task.

A Simple Example of Task Modeling in MONACO_T

To take care of a patient, in an intensive care unit, we have at least three persons : a doctor, a nurse and a technician laboratory. The static of the task to perform is illustrated by the [Fig. 1] :

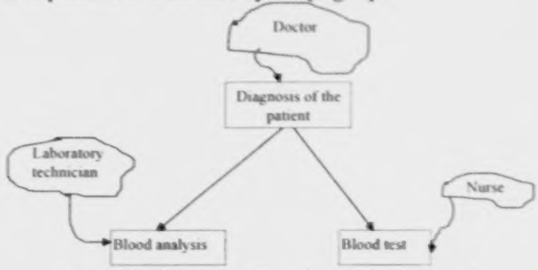


Figure 1 : decomposition of a cooperative task

The dynamic of the task and his sub-task is defined by the following rules embedded in each node of the static tree.

dynamic of the sub-task Diagnosis
Rules of activation : *Diagnosis* state = **initial**.
Rules of realization : *Diagnosis* state = **activated** AND *Blood analysis* state = **ended**.
Rules of termination : *Diagnosis* state = **realized**.
dynamic of the sub-task Blood test
Rules of activation : *Blood test* state = **initial** AND *Diagnosis* state = **activated**
Rules of realization : *Blood test* state = **activated**.
Rules of termination : *Blood test* state = **realized**.
dynamic of the sub-task Blood analysis
Rules of activation : *Blood analysis* state = **initial** AND *Blood test* state = **ended**
Rules of realization : *Blood analysis* state = **activated**.
Rules of termination : *Blood analysis* state = **realized**.

Architecture of The Learning System

In the context of teaching a cooperative task, a learning system must be able to teach the knowledge related to each role of the task. It has also to teach the team how to coordinate themselves in order to well-perform collectively. The architecture [Fig. 2] we propose relies on INTERNET. It is structured around a tutor server which maintains the

BEST COPY AVAILABLE

communication between all the tutor clients. These ones are implemented as agents and they discuss with their learner like a real tutor would do.

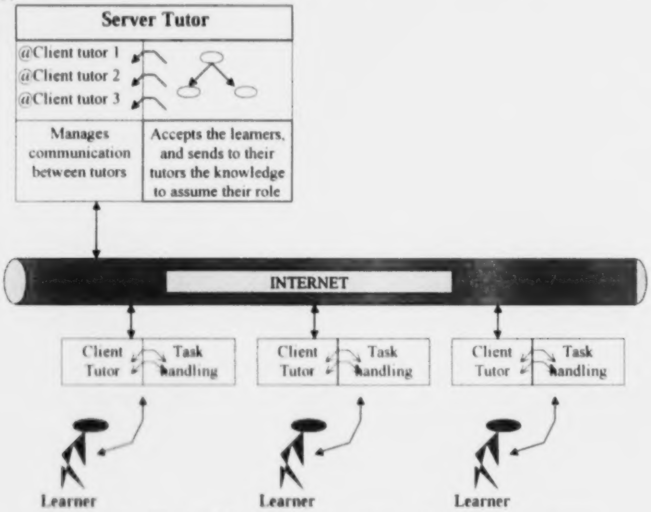

Server Tutor
@Client tutor 1
@Client tutor 2
@Client tutor 3
Manages communication between tutors
Accepts the learners, and sends to their tutors the knowledge to assume their role
INTERNET
Client Tutor
Task handling
Client Tutor
Task handling
Client Tutor
Task handling
Learner
Learner
Learner

Figure 2 : The architecture of a cooperative learning system

The Behavior of the System

The system only implements one strategy of learning : the learning by doing. This strategy means that when a has to learn a task, the system sets all the parameters of the task and lets the learners try to execute it. If one of them has a problem to achieve his role, his tutor gives him the help he requires. In this case, the tutor may need an information from another role. He then contacts the concerned tutors and communicates with them to construct the best explanation for the learner.

The Architecture of Tutor Client

The architecture of a client tutor agent should contains a knowledge and an inference engine which allows him to reason and give all the help needed by a learner doing his task. This tutor should also have the possibility to communicate with his peers to build explanations with the contribution of the whole team. The architecture [Fig. 3] we propose for the tutor contains four parts : the long time memory, the working memory, the reasoning module and the dialogue module.

The Long Time and the Working Memory

The long time memory contains all the knowledge which helps the tutor to produce all the explanations needed by the learner, and to communicate with his peers. The information in the long time memory is divided in three blocks:

- The rules of activation, realization and termination which allows the tutor to manage the sub-tasks of his role,
- The methods which allow the tutor to simulate the activity related to his role,
- The addresses of the other tutors.

The working memory contains the whole states of the cooperative task. It also gathers all the data representing the evolution of the task execution. It can be consulted by all tutor agents. Each of them has also the exclusive right to update the subset of the working memory which is dedicated to him.

BEST COPY AVAILABLE

The Reasoning Module

According to some recent work in agents domain [Frasson et al, 96], [Ferber, 94], [Terveen, 95], agents can do reasoning according to four levels : reactive, cognitive, social and learning. In our architecture, we have implemented the reactive and the social levels. We have also created a new kind of reasoning : the cooperative level. All these levels are used by the system to answer several questions like: how to realize an action ?, why an action can't be realized ?, why do I need to realize this action ? We will use the last question to illustrate this behavior.

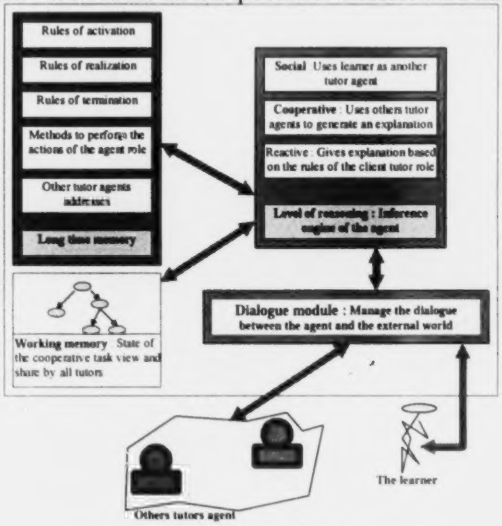

Rules of activation
Rules of realization
Rules of termination
Methods to perform the actions of the agent role
Other tutor agents addresses
Long time memory
Social Uses learner as another tutor agent
Cooperative : Uses others tutor agents to generate an explanation
Reactive : Gives explanation based on the rules of the client tutor role
Level of reasoning : Inference engine of the agent
Working memory State of the cooperative task view and share by all tutors
Dialogue module : Manage the dialogue between the agent and the external world
The learner
Others tutors agent

Figure 3 : Architecture of an client tutor

Reactive Level of Reasoning

This level of reasoning allows a client tutor to respond instantaneously to learner questions. This reaction is based on the information in the long time memory. In the example given previously, where a learner plays the role of a doctor, and the cooperative task is in its initial state, the learner can ask: What do I need to realize the diagnosis action ? In this case the doctor response will be: You have to **activate** the *diagnosis* sub-task and the laboratory technician have to **end** the sub-task *blood analysis.*

Cooperative Level of Reasoning

In this level of reasoning, when the tutor has to give an explanation to the learner, he can ask for the cooperation of the other tutor agents if it is useful, in order to generate the best explanation which takes into consideration the entire cooperative task. The following example demonstrates the way the tutor is using to produce an explanation.

A learner is playing the doctor's role, and the cooperative task in it's initial state. We also assume that DL=doctor learner, DT=doctor tutor, NL=nurse learner, NT=nurse tutor, TL=technician learner and, TT=technician tutor.

DL asks to DT : What do I need to **realize** the*Diagnosis* sub-task ?

The DT analyzes the question and see that DL has to **activate** the *Diagnosis* sub-task and that TL has to **end** the *Blood analysis.* In order to know why the *Blood analysis* is not **finished** DT is going to ask TT.

DT asks to TT : What does TL need to **end** *Blood analysis* ?

TT asks to NT : What does NL need to **end** *Blood test* ?

NT asks to DT : What does DL need to **activate** *Diagnosis* ?

Because nothing prevents DL to **activate** *Diagnosis*, The team of tutors will began to construct the explanation.

DT responds to NT : DL needs nothing

NT responds to TT : Response of DT + "DL has to **activate** *Diagnosis"*

TT responds to DT : Response of NT + "NL has to **end** *Blood test"*

BEST COPY AVAILABLE

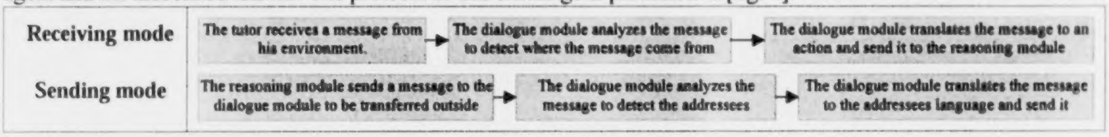
DT responds to DL : Response of TT + "TL has to **end** *Blood analysis*".
The final response contains all the preconditions which have to be satisfied before DL **realizes** *Diagnosis*.

Social Level of Reasoning

In the social level of reasoning our system uses a team of tutors augmented with all the learners to construct cooperatively an explanation. This characteristic is an original feature and gives to the learning system the ability to interact with a leaner without being asked.
The behavior of this level is the same as the behavior of the cooperative level except when a tutor detects that a learner can do some action which unfreeze another learner. Instead of stopping the propagation, the tutor wakes up the learner by sending him a question. This question can initiate a discussion between him and his learner in order to teach this one that his action is very important since they are executing a cooperative task.

The Dialogue Module

This module allows the tutor agent to reorganize the explanations produced by the reasoning module. This reorganization depends on who is talking with the agent. It also manages exchanges between two agents or between an agent and his associated learner. The protocol of this exchange is presented in [fig. 4].



Receiving mode	The tutor receives a message from his environment.	The dialogue module analyzes the message to detect where the message come from	The dialogue module translates the message to an action and send it to the reasoning module
Sending mode	The reasoning module sends a message to the dialogue module to be transferred outside	The dialogue module analyzes the message to detect the addressees	The dialogue module translates the message to the addressees language and send it

Figure 4 : Protocol of dialogue module

The quality of service provided by the net is used by this module to manage the explanations. If the quality of service is bad, this management consists in giving an explanation at the reactive level even if the learner needs it at the cooperative or social level. The choice allows the tutor to answer quickly to the learner.

Implementation

The implementation of this system is in the JAVA language. The server tutor is a JAVA application, and the clients tutors are JAVA applets. When a learner wants to be connected to the system, he uses any HTML browser with the URL of the client tutor applet [Fig. 5].

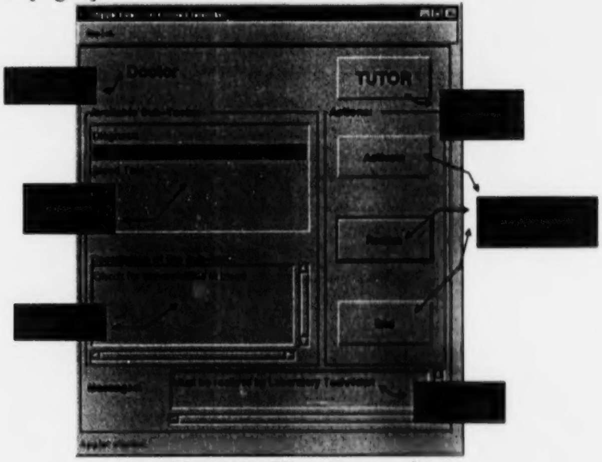


Figure 5 : Client tutor applet

BEST COPY AVAILABLE

When an agent tutor wants to send a message to a peer, this message is transmitted to the server tutor before being sent to the addressees. So the architecture of communication is a star net centered on the server tutor. The client tutor and the server tutor communicate by using sockets, serialization and RMI(Remote Method Invocation) objects in JAVA.

Conclusion

The system we built gives the feeling of having only one virtual team advisor. All the learners have the same goal, the one of achieving the cooperative task. The members of the group depend on each other to accomplish a shared goal or task. Without the participation of one member, the group is not able to reach the desired goal. Because each member of the group is held accountable for the goal of the team, our system has the capability to learn social skills. These skills help the learners to achieve efficiently their role in the society.

At present time, we have experimented our system in the case of a small cooperative task. We are planning to use this system to train government employees in manipulating a cooperative document like a birth certificate or a marriage certificate.

Bibliography

[Aimeur et Frasson, 96] Aimeur, E. et Frasson, C. (1996). Analyzing a New Learning Strategy According to Different Knowledge Levels *In Computers Educ volume 27 Numbers 2, pp 115-127.*

[Brézillon 92] Brézillon, P.(1992). *Explication et coopération:Une contribution bibliographique au travail du groupe COOP.* Rapport No 92/31

[Card, Moran, & Newell 83] Card, S., Moran, T. & Newell, A. (1983). The Psychology of Human-Computer Interaction, Hillsdale, NJ : Erlbaum.

[Chan et Baskin 90] Chan T.W. et Baskin A.B.(1990). Learning companion systems. *InIntelligent Tutoring Systems : At the Crossroad of Artificial Intelligence and education (Edited by Frasson C. And Gauthier G.) Chap. 1. Ablex, N.J.*

[Djamen 95] Djamen J.Y. (1995). Une architecture de STI pour l'analyse du raisonnement d'un apprenant. Thèse de Ph.D., DIRO, Université de Montréal, Montréal, CANADA, Juin 1995.

[Ellis, Gibbs et Rein 91] Ellis, C., Gibbs, S.J., et Rein, G.L. (1991). Groupware: Some issues and experiences. *CACM*, 34(1), pp. 38-58.

[Frasson, Mengelle, Aimeur et Gouarderes 96] Frasson C., Mengelle T., Aimeur E., Gouarderes G, (1996). An Actor-Based Architecture for Intelligent Tutoring Systems, In *The Proceeding of Thirth international Conference of Intelligent Tutoring Systems*, June, Montreal, Canada.

[Nkambou et Gauthier 96] Nkambou, R., Gauthier, G., (1996) Un modèle de représentation de connaissances dans les STI. In Proceeding of 3th International Conference on Intelligent Tutoring System ITS96. Springer-Verlag.

[Schmidt et Bannon 92] Schmidt, K. et Bannon, L. (1992). Taking CSCW Seriously. *In Computer Supported Cooperative Work Journal, 1,1*, pp. 7-40.

[Tadié 96] Tadié, G.S., (1996) Coopération et Système Tutoriel Intelligent. Rapport de l'examen prédoctoral oral. Département d'Informatique et de Recherche Opérationnelle. Université de Montréal.

[Tadie, Lefebvre et Frasson 96] Tadie, S., Lefebvre, B. et Frasson, C. (1996). MONACO_Tâche: un modèle à nature coopérative pour la modélisation de tâche dans un STI. In *The Proceeding of Thirth international Conference of Intelligent Tutoring Systems*, June, Montreal, Canada.

[Terveen 95] Terveen, L.G.(1995). Overview of human-computer collaboration. *In Knowledge-Based Systems volume 8 Numbers 2-3.*

Acknowledgment

This work is a part of TLCE(Tele Learning Center of Excellence) project of CANADA.

A Foot In Both Camps: Interventions in National and Global Regulatory Processes by Nation-based Internet Organisations

Jenny Shearer
Computer Science Department and Political Studies Department, University of Auckland, New Zealand.
jshearer @cs.auckland.ac.nz

Abstract: The Internet Society of New Zealand provides a case study for discussion of how national Internet organisations may optimise the future of the Internet and protect its developing infrastructure, while maintaining the interests of their nation states. The role of the national Internet organisaton should be to mediate between the interests of the Internet community and the interests of the state, in providing a political communications channel and liaison with global counterparts, and in proposing innovative solutions to conflict. Key issues of Internet infrastructure are outlined with an analysis of optimal responses within the perameters described.

Introduction

The Internet Society of New Zealand (http://www.isocnz.org.nz) works for a technologically advanced but small nation in a context of international regulation and market forces. [Hicks 1996] Policy initiatives in the areas of Domain Name System management, cryptography and Public Key Encryption (PKE) Certification Authority issues, a Code of Practice for Internet Service Providers, copyright regulation, and a policy link with the USACM, aim to maintain the Internet as a free and open system, and solve problems posed by political pressure towards legislation and regulation perceived as inappropriate by the Internet community.

Most national governments have expressed support for the Internet and its opportunities for improved global communications and trade opportunities. However, nation states are also supporting their own interests, broadly represented as economic welfare, national security, and the preservation of the political power of the state. The resulting conflict of interest means governments are contributing to regulatory activity which may significantly restrict development of the Internet. Internationally, the effort to hammer out global accords which will protect various commercial and national interests has resulted in unsatisfactory proposals for cryptography use and copyright protection. It is indeed evident that the advent of the Internet has profound implications for the traditional relationship between citizens and state. Traditionally, the Hobbesian argument that the state protects its citizens and in turn, the citizens accept the jurisdiction of the state, becomes less relevant in a global communications forum and marketplace. In the open "cyberspace" of the Internet, nations and their demands in terms of taxes and observance of cultural mores may readily be ignored, in the absence of penalties.

Rossnagel proposes a part solution:
"When the democratic constitutional state can no longer reliably protect its citizens in the new social space of the networks, in compensation it must enable them to protect themselves." [Rossnagel 1997]

"Some of these measures - for example the encryption program PGP - can be used without any advance concession. The state only has to abstain from impeding regulations. Others - such as digital signatures - depend on an infrastructure that allows the individual to use these protective measures. The citizen of the information society still depends on infrastructural prerequisites. But there is a fundamental difference in whether the individual can decide about using self-controlled protective measures himself, or the state or another large organisation offering protection that he cannot influence."

"In order to protect and preserve the "old" goals of freedom and self-determination in the "new" social space of the networks, law must permit and support new technologies." [Rossnagel 1997]

"The Internet community" consists of Internet users who share beliefs, such as that the Internet should maintain freedoms of speech, of access to information in the global public domain, and that its development in technical and societal terms should not be hindered by the demands of markets, or individual nations. It flirts with concepts of global citizenship. [Shearer 1996]

National Internet societies or organisations with an interest in the Internet are evolving into a position of influence in both worlds, those of national governments and the Internet community. It is their role to guide national governments into a period of power transfer towards global settlement of outstanding Internet issues, without loss of integrity to the nation state. They have an important role in political interventions, interpreting technical issues of Internet operation to their public, creating strategies to ameliorate public concern over pornography, terrorism and the like, and creating strong Internet infrastructure such as public key encryption Certification Authorities, domain name systems, and other items, in the event of market failure to carry out these tasks. Globally, they have the task of building links with other national organisations to create a rapid and effective political response by the Internet community to attempts to impose inappropriate regulation. To this end, the Internet Society of New Zealand (represented by the author as a member of the Society's Council) is drawing up an agreement with the USACM public policy committee to facilitate collaboration between the two organisations on global policy matters. It is hoped the agreement may be a blueprint for co-ordinated response by a number of national Internet organisations in defending the Internet against damage by inappropriate regulation. This policy collaboration may be seen as complementing international committee initatives such as that of the Internet Society (ISOC) which aims to bring together users, service providers, standards bodies and government organisations with the intent of setting global policies for the Internet. In the national environment, the Internet Society of New Zealand has set up a company to run the domain name system. This has assured a consistent domain name infrastructure, while allowing for limited competition in administration of second level domains.

1 Cryptography Policy

Cryptography is viewed by the Internet community as critical to the use of the Internet as an open political and public forum, and for use as an essential security measure for the development of electronic financial transactions. A large number of national governments view the widespread use of "strong" cryptography as a potential threat to national security, if used by terrorists or drug dealers, and as a means for citizens to evade paying tax. This last concern has major implications for the ability of national governments to remain in power, if their source of income is cut off. Most of these concerns may be challenged, and should be weighed against the political implications of widespread Government surveillance resulting from key escrow or key security measures, and the right of citizens to privacy of communications. [Shearer and Gutmann 1996].

The Internet Society of New Zealand has written to the New Zealand Government urging that the Government enter OECD discussions to put forward a position on cryptography that supports the free and open use of cryptography within countries and internationally. New Zealand's stable democracy, high uptake of Internet services, its relative independence in policy decisions, and its significant expertise in areas such as computer science, cryptography, and communications policy, means it is possible to present an optimal environment in terms of cryptography use.

New Zealand has no formal restrictions on the use of cryptography within the country, and cryptography developers have made major advances in this field. New Zealand is a participant in the Wassenaar Arrangement, and it is a cause for concern to the Society how this arrangement has been interpreted by Government departments. (The Wassenaar Arrangement on Export Controls for Conventional Arms and Dual-Use Goods and Technologies states that "national discretion" may be used in its application. There is no requirement for governments to impose controls on cryptographic items, and indeed many countries which are signatory to the agreement have no export controls on cryptography, or have token controls which are not enforced.) Few export licences have been granted for current exports of cryptography products from New Zealand; some of those granted have been to the United States. The Internet Society believes the cryptographic export licencing structure to be unjustified. The Society supports the Government's initiative in moving to set up a public service Public Key Encryption registry, which will enable public service institutions to protect their confidential communications with strong cryptography. This service will be set up by the Government Communications Security Bureau, and given that this organisation is responsible for Government surveillance, the Society has requested an assurance from the Government that key escrow/key security measures will not be put in place.

The Internet Society is investigating means of establishing a private sector PKE registry, to provide support for commercial and individual cryptography use. In part, this initiative aims at the establishment of free use of "strong cryptography" to demonstrate that citizens will not use this to destabilise the Government, and to establish a public view that cryptography regulation is unnecessary and an infringement of their rights.The Internet Society believes free use of cryptography is necessary for the Internet to develop to its full potential in commercial and societal terms, and that interventions by national governments to impose cryptography regulation or key escrow/security systems will result in damage to the global infrastructure. Such interventions are not justified by concerns about perceived problems of individual national security. The US, for example, has transferred encryption items from its Munitions List to the Commerce Control List, enabling export from the US of 56 -bit encryption items, with the proviso that a key recovery infrastructure will be put in place. The USACM public policy committee has stated in a letter to the US Commerce Department that the policy will hinder US and international research efforts. Further;

"Key recovery products have not yet been subject to the vigorous testing necessary for a proposed standard and there is little understanding of how such a system would operate and what controls would be needed to ensure that it remained secure.
.."We believe the Commerce Department should not promulgate regulations which prohibit US research and development from responding to market demands and limit the ability of Americans using new on-line services to protect their privacy."

The US Special Envoy for Cryptography, David Aaron, commented at an RSA Data Security Conference that "everyone involved with the encryption issue, whatever their views, recognises that international reaction will determine the success or failure of their particular approach." [Aaron 1997]

In similar vein, a working document for a September 1996 meeting of the AD Hoc Group of Experts on Cryptography Policy Guidelines for the OECD warns:

"Efforts by a single national government to regulate the use of cryptography in ways that are incompatible with other national governments pose a serious risk that the regulating government's policies will be ineffective.
"While recognising that a state's sovereign responsibility to protect public safety and national security may require it to take unilateral action disparate national policies will also impair the development of the GII/GIS." [OECD 1996]

The attempt by the US to impose a global standard of key recovery systems may fail if the Internet community creates an international climate where this is unacceptable. Internet organisations may play an important part by creating a public awareness of the problems, and by lobbying governments on the issue.

2 Code of Practice

A Code of Practice for Internet Service Providers [Code of Practice 1996] has been developed under the auspices of the Internet Society of New Zealand. The Code has been in essence a response to a public demand for something to be seen to be done about the problem of pornography, with a background threat of major legislative intervention. Shortly after the introduction of the Communications Decency Bill in the US in 1995, a New Zealand MP, Trevor Rogers, put forward a private member's bill in the New Zealand Parliament, the New Zealand Technology and Crimes Reform Bill. The Bill proposed to make Internet carriers legally responsible for " objectionable" pornography [Films, Videos, and Publications Classification Act] transferred over their lines. The Bill was referred for select committee hearings, where the Internet community raised major objections. During the discussions, Waikato and Victoria Universities, who were responsible at that stage for all Internet connection, announced they would review it ("pull the plug") if the Bill went through. A Justice Department review found the legislation could bring about a "chilling" of public expression. In select committee hearings, warnings were given by Internet advocates of major economic and social disadvantage to New Zealanders if measures proposed by the Bill were implemented.

In 1996, New Zealand Minister of Communications, Maurice Williamson, suggested the Internet Society of New Zealand develop a Code of Practice for Internet Service Providers (ISPs). The Internet Society, in consultation with ISPs began the process of developing a Code of Practice which would address the question of objectionable material being transferred via Internet, and set up operating standards for ISPs. This is viewed by

Society Council members as an opportunity to address issues of concern to the public in a substantive way, while forestalling future attempts to impose legislation which might be damaging to freedom of speech on the Internet, as well as placing unacceptable burdens on Internet service providers. The Code of Practice, as a developing document, may also be amended to take in issues of copyright as they affect ISPs. The Technology and Crimes Reform Bill was officially dropped by the New Zealand Parliament in July, 1997.

The Films, Videos and Publications Classification Act has been used to prosecute people distributing objectionable material via bulletin boards based in New Zealand. The Internet Society has also mediated in a situation where New Zealand Department of Internal Affairs inspectors in 1996 approached individual ISPs to try and track users who were allegedly using the Net to exchange objectionable material. In a press release, the Internet Society chairperson, Jim Higgins commented:

"We discussed (with the Internal Affairs Department) the fact that much of the existing legislation, implemented before the Internet became the ubiquitous telecommunications platform it now is, is quite unsuited to handling these problem and places innocent parties such as the ISPs at risk of prosecution for 'storing offensive material'.

"Internal Affairs has agreed to take a pragmatic approach and focus on the end users of objectionable material, rather than ISPs who unknowingly and unwillingly pass it through their computers."

"ISOCNZ is currently well down the track with developing a Code of Practice for ISPs, and we see that this, together with a revamp of legislation to protect the ISPs in the same way that NZ Post and the telephone companies are protected, should help considerably."

The voluntary Code of Practice draft includes the following aims: to impose and regulate industry standards, to protect rights of access and free speech, and to ensure that information and procedures are in place for the protection of minors from accessing objectionable material over the Internet. Procedures for dealing with complaints are to be developed.

Code of Practice members are required to support the tagging in of URLS and other content, related to educational/childrens' content. Adult services hosted should be classified by PICS [PICS, 1997] or other common systems, should be segregated, subject to pin type security and with identifiable signatures, should be accompanied by on-screen warnings, managed by subscription enrolments to exclude under-age subscribers and finally, members should support adoption of a system of tagging in URL's related to adult services. In the area of electronic commerce, standards for sale transactions on the Internet are set up, including specifics of information to the customer.Further sections deal with customer education and dispute resolution. A footnote comments that Discussion during development of the Code was inconclusive on the issue of setting up "family accounts" or "family safe areas" to be administered by ISP's, due to the current difficulty in assuring the exclusion of undesirable material. A footnote comments that ISPs would have to offer "best effort", rather than accepting absolute liability for content.

The "traditional" Internet philosophy of upholding complete freedom of speech has resulted in a fragmented response by the Internet community to public concerns, well aired in the traditional media, about pornography and "bomb recipes".However, national Internet societies are well-placed to educate their populations and develop nationally-appropriate solutions. The proposals may include implementing leading-edge screening or network technology and developing a policy framework :

- recognising the traditional claim of sovereign states to organise their own affairs in terms of the level of "offensive" material citizens are prepared to tolerate, and developing technological means, to allow such national censorship to be carried out.

- creating a powerful on-line presence and political lobbying force to examine and counter exaggerated or unbalanced public discussion or legislative activity, in order to uphold the principle of freedom of speech in the Internet.

The above proposals are not mutually exclusive, but are a recognition that, in order to move the Internet forward, some ground must be given. This is in order to prevent the public forum potential of the Internet foundering, for example, on a traditional expectation in many societies including the US, that children be protected from viewing offensive material.

3 Copyright

The Internet Society of New Zealand wrote to the New Zealand Government in opposition to copyright clauses of the WIPO treaty, raising issues specific to the operation of the Internet in New Zealand, along with issues regarding browsing and databases raised by the USACM. A reply noting the Society's concerns was received prior to the New Zealand delegation attending the WIPO meetings in Geneva in December 1996. The New Zealand Ministry of commerce reported back on the WIPO Treaties to interested groups at a meeting on February 24, and invited these groups to submit their views on national policy positions on copyright. The Society will maintain an interest in issues relating to databases which remain to be negotiated by WIPO.

Collaboration with the USACM public policy committee enabled the Internet Society of New Zealand Council to utilise the skills of USACM analysts to quickly develop a policy position on copyright , when news of the WIPO Treaty provisions first broke late in 1996, and were considered by Internet commentators to be potentially damaging to the Internet. In a letter to New Zealand Government Ministers, the Society said:

"The Internet Society of New Zealand recognises the need to protect investments made in large data collections. However we believe the draft Treaty in its present form could prevent the Internet community from accessing much information which has traditionally been in the public domain, and could also greatly hinder the ability of Internet users, from schoolchildren, scientists, and commercial users, to pursue their interests and work.

For reasons detailed in this letter, we urge the New Zealand Government to reevaluate several key areas of the draft Treaty, and we advise that it would not be in the interests of New Zealanders to vote for it until major revision and discussion has taken place.

Key issues are that of how "fair use" of protected information would work, that is, the right of people to look at copyrighted work, and extract and comment on sections of it, without being in breach of copyright. The issue of "temporary" copies of work briefly looked at by a computer user which would be considered a copy for copyright purposes under the draft Treaty, would impact on the important "browsing" nature of current Internet use. Though "fair use" is protected under New Zealand copyright regulations, New Zealand regulations in respect of this and other issues would have to closely follow a Treaty agreement, meaning that existing New Zealand regulations are likely to be overruled.

Under the Treaty, those considered to be the "owners" of a set of data, even if it was weather reports, sports results, financial data, or other information formerly deemed public information, would be able to limit access to it. Libraries, for example, could suffer major adverse effects. Perpetual protection granted to databases under the Treaty is extreme given the time limits conferred by traditional intellectual property laws. The issue of inclusion of importation rights requires further discussion.

In short, the provisions of the Treaty appear to have been devised to benefit commercial publishing interests at the expense of maximising the present and future benefits of databases to the global community, of which New Zealanders are a part." [WIPO letter, 1996]

4 Conclusion

Policymaking in the new Internet environment requires national organisations to monitor developments across the fields of technology, politics (both national and international), commerce, and culture. In developing policy solutions which reconcile the interests of the Internet community with those of the nation state, these organisations require flexibility and the ability to react quickly. Policy collaboration with other national organisations is a way for national organisations to add to their own policy resorces, and to respond in a co-ordinated way to issues of major concern to the Internet community, such as the proposed WIPO treaties on copyright. As mediators and leaders of public opinion, and as keepers of Internet infrastructures, the national Internet organisations have a significant role to play in the development of the Internet.

The Internet Society of New Zealand is an independent national voluntary body, with a Council elected by a membership representing Internet interests across the board in New Zealand. Policymaking reflects the objectives of the Society in protecting and promoting the Internet to enable the full potential of the technology in commercial, societal, and global terms to be achieved. Perceived rights to privacy of communications, free and open operation of the Internet, universal access, and effective infrastructure, inform policy development by the Society. A broad consensus on these objectives is achieved within the New Zealand Internet community, and is reflected in Government policy.

5 References

[Hicks 1996] Hicks, R. (1996). *The Internet Society of New Zealand: Roles, Goals, and Ambitions*. ACM StandardView Vol.4, No.3, September 1996.

[Rossnagel 1997] Rossnagel, A. (1997). *Global Networks: State's Impotence, Citizens' Self-Protection*. Zeitschrift fuer Rechtspolitik 1997, pg26 ff.

[Shearer 1996] Shearer, J. (1996). *One Net, One World. Global Citizenship and the Internet*. Journal of Universal Computer Science, 2 (12) pg 842.

[Shearer and Gutmann, 1996] Shearer, J. & Gutmann, P. (1996) *Government, Cryptography, and the Right to Privacy*. Journal of Universal Computer Science, 2 (3) 113-137.

[Aaron 1997] Aaron, D. (1997) *Speech, January 28, to RSA Data Security Conference*, San Francisco. <http://www.bxa.doc.gov/aaron.htm>

[OECD 1996] OECD (1996) *Ad hoc group of experts on Cryptography policy guidelines, Working Document for 26-27 September 1996 Meeting*. <http://193.154.75.3/Netzeil/OECD/oecd_05.html>

[PICS 1997] (1997) *PICS* <http://www.w3.org/pub/WWW/PICS>

[Films, Videos, and Publication Classifications Act] *Films, Videos, and Publications Classification Act*, Published by the New Zealand Government Printer.

[Code of Practice 1996] Internet Society of New Zealand. (1996). *The Internet Service Providers Code of Practice, Draft 1.1*, November, 1996. <http://www.isocnz.org.nz>

[WIPO letter 1996] Internet Socity of New Zealand . (1996). *Letter from the Internet Society of New Zealand to Ministers of the New Zealand Government, on WIPO treaty negotiations on Copyright*. <http://www.isocnz.org.nz>

BEST COPY AVAILABLE

A Survey of Web Information Systems

Alberto Silva
Alberto.Silva@inesc.pt
IST/INESC

M. Mira da Silva
mms@dmat.uevora.pt
University of Evora

José Delgado
Jose.Delgado@inesc.pt
IST/INESC

Abstract:
The main objective of this paper is to survey, in a schematic and generic way, the major models and approaches concerning the design, construction and execution of Web information systems. We identify three different approaches – server-centric, client-centric and distributed applications – and discuss the major strengths and weaknesses of each approach in a Web context. Finally, we present our view of those models and technologies that are needed for building the kind of distributed, dynamic information systems that will exist in the near future.

1. Introduction

The Web grew fast from a distributed hypermedia system, with basic navigation and information retrieval capabilities, to what we call an "umbrella system" of several and distinct applications (also called information systems or simply IS) eventually accessible at a world wide scale.

These days, Web Information System (WIS) technology involves many efforts, interests and investments all over the world, spread by the main research and academic centers as well as the major IT companies. Due to this highly dynamic situation, it *is not* the aim of this paper to describe specific tools, environments, or products concerning WIS development because they would be quickly outdated by new generations of (yet more competitive and complex) products.

The main objective of this paper *is* to describe, in a schematic (see schema in annex) and generic way, the *main models or approaches concerning WIS design, construction and operation*. Consequently, we discuss first the major strengths and weaknesses of the current Web technology and will also depict what models and technology of the distributed and dynamic WIS should exist in the future.

The paper is a result of our real-world experience on designing and developing large distributed applications, as well as our research work. More recently, the first author has pursued research on developing distributed information systems on the Internet [SBD95,SAD96,SMdSD97], the second on higher-order distribution mechanisms for persistent programming languages [MdS96,MdSS97], and the third on visual development environments and intelligent building systems [DSCOB95].

The paper is organized in six sections. The next section presents the server-centric WIS approach (i.e., the class of WIS in which activity is concentrated on the server machine) and section 3 presents the client-centric WIS approach. In section 4 we describe a new approach to WIS based on distributed infrastructures, where the activity is effectively divided between client and server machines. Finally, section 5 summarizes and compares all approaches, including the future models for WIS.

2. Server-Centric WIS

We define as server-centric WIS those information systems based on the Web that concentrate their activity on the server side. In this survey we will describe the common gateway interface, server side includes and proprietary server interfaces like ISAPI. The section ends with a comparison between these variants of the basic server-centric WIS model.

2.1 Common Gateway interface (CGI)

The first server-centric WIS were based on the CGI (*Common Gateway Interface*) [Rob96]. This interface consists in a generic and simple specification that defines a communication interface between a Web server and any specific process that runs in the same computer architecture.

CGI is a flexible and extensible way to add functionality to a Web server, such as: data base access, data translations, protocol conversions, and so on. CGI programs are typically written in a well-know language (such as C or C++) or in a simple, powerful script language (such as Perl).

After the construction of the first real WIS based on the CGI, some variants happened, either for simplicity reasons (Server Side Includes mechanism, see section 2.2) or due to performance stress (Web server proprietary APIs, see section 2.3). Nevertheless, all server-centric WIS share the following common issues:

- WIS are called "short-life" processes because Web servers are stateless and the HTTP protocol is connectionless. As such, a CGI process is be responsible for maintaining state (persistent data) between successive user interactions. There are some known techniques for simulating state between different interactions (accesses) in the same session, for example: using: URL information (e.g., *QueryString* and *Path Info* components); visible or hidden form's fields; cookies; etc.

- The majority of these applications are HTML document generators, which may vary between trivial or very complex. This means that the main function of any WIS is to parse a set of data sent by the Web client component (the browser) and produce at run-time a convenient response, which is typically in HTML format [SBD95].

Figure 1 presents the computational model of WIS based on the CGI approach. Specific components of the application are presented with emphasis (gray color): the CGI process and the purposely non identified "??" component. This "??" block is usually a more specialized process with which the CGI process would have to communicate, for example, a file server, a data base management system, a geographic information system, or any other more specific process. The other blocks – Web client and server – are generic, just used as support components (consequently, in this paper, we don't give them any special relevance).

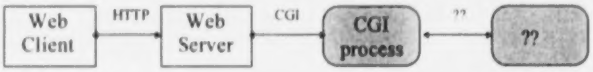


Figure 1: WIS computational model based on the CGI approach.

2.2 Server Side Includes (SSI)

The Server Side Includes (SSI) mechanism was originally introduced in the NCSA's Web server [NCSA95a] to facilitate the inclusion of simple extensions to the Web server.

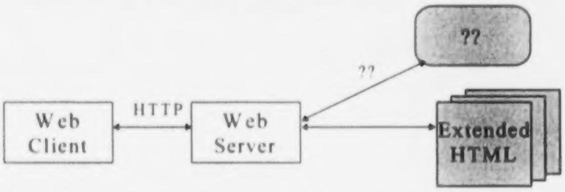


Figure 2: WIS computational model based on the SSI approach.

Figure 2 depicts its generic functionality. The application is basically composed by a given set of HTML files with some extended (non standard) elements called *tags*. When the Web server receives a request for some extended HTML document, it parses the corresponding file and translates every tag into a set of "standard" HTML elements. To translate the extended elements, the Web server may eventually need to invoke external processes (such as CGI programs, DBMS, etc.) which are represented by the "??" block with a "??" interface.

SSI was originally conceived for the NCSA Web server. However, it has inspired some equivalent mechanisms, although more flexible and skillful, such as Microsoft's Internet Database Connector [Mic96a] (designed for accessing databases via ODBC) and WebQuest's SSI++ [Ques96]. Nevertheless, its main goal was just to enable the development of simple WIS in an easy and fast way – even without the need for any programming language.

2.3 Proprietary APIs

In this approach, WIS use an application programming interface (API) proprietary to a specific Web server such as those from Netscape or Microsoft. Unlike both CGI and SSI, this permits to write and load new programs in the same memory space of the Web server itself. Sun's Servlets (Java components executed on the server-side) technology may be also classified in these approach.

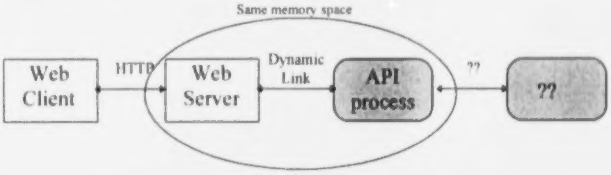


Figure 3: WIS computational model based on the proprietary Web server API approach.

Figure 3 depicts the computational model corresponding to this approach. The load (also called binding or linking) operation is executed once at server boot time or on demand each time they are invoked by a client request. This approach doesn't solve the "short-life cycle" problem presented by all server-centric WIS, but eliminates, for each access, a new process creation. The only requirement is the support to dynamic linking of executable modules by the operating system, an almost standard feature in modern OS like Windows 95.

Currently, almost all well-known Web servers provide their own proprietary API, for example: Netscape's NSAPI (*Netscape API*), Microsoft's ISAPI (*Internet Server API*), and Oracle's WRBAPI (*Web Request*

BEST COPY AVAILABLE

Broker API). However, since it is impossible to support a large number of APIs, most software development tools such as Borland's Delphi choose to support only one or two of these, typically Microsoft's ISAPI (when based on Windows) and/or NSAPI (since Netscape is more popular in the commercial UNIX world).

2.4 Comparative Analysis

In this section we summarize and compare the three approaches: CGI, SSI and API. Obviously, the classification presented in Table 1 represent only high-level classification and are not intended to be used as a means to decide which approach is the best.

Analysis criteria	CGI	SSI	API
Application performance	☹	😐	☺
Application scalability	😐	☹	😐
Application safety	☺	☹	☹
Application flexibility	☺	☹	☺
Easy development	😐	☺	☹
Application maintenance	😐	☺	☹
Vendor-neutral	☺	😐	☹

Table 1: **Comparative analysis of server-centric WIS approaches.**

The aim of this analysis is to enable a deep reflection of the strengths and weaknesses of the different approaches, and as a result the following criteria are analyzed: runtime absolute performance; scalability (the ability to support a large number of client requests); safety (the consequences to the Web server if a malfunction occurs in the application); flexibility and versatility; easy development and technical skills required; application management and maintenance; and open (vendor-neutral) against proprietary solution.

WIS based on CGI

These have low-level performance due to their deficient management of processes and memory. However, they can support some degree of scalability via replication of CGI processes on different machines. (This is achieved by a special load balancing algorithm in the code of each CGI application.)

This approach presents a high level of safety because applications are external to Web servers, meaning that if the process crashes it doesn't imply a corresponding Web server crash.

These CGI applications are very flexible and even complex because they can be developed independently of the Web server using any conventional language (C, C++, Perl, Basic, Java) and development environment (Visual Basic, Delphi, PowerBuilder).

The development and test cycle is moderated. It all depends on the capabilities of the chosen language and development environment and also depends on the existence (or not) of specialized software libraries supporting the CGI particularities. These libraries typically give support for: state maintenance; HTML code generation, access to a DBMS; etc. Technical skill requirements, application management and maintenance is medium or high depending on the very same characteristics.

Due to its simplicity and support, it was adopted by the generality of Web servers. CGI then became, and still is, the standard *de facto* for building WIS.

WIS based on SSI.

This approach is basically a substitution mechanism, in which tags (non-standard HTML elements) are converted at "request-time" to a corresponding set of standard HTML elements. Examples include: the result of a system call (e.g., current time) or the result of a SQL query.

SSI presents a medium/high performance because the server has to parse the page and substitute all tags on it every time that page is requested. As a result, scalability is not well supported.

Its main drawback consists of the limited capabilities concerning the development of flexible and complex applications, which are impossible without a computationally complete programming language like C.

If the application crashes when translating one of these tags, that implies the crash of the corresponding Web server. Also, it is difficult to develop an independent-vendor application because each vendor defines its own set of tags. Nevertheless, its corresponding development is easy and fast for very simple applications.

WIS based on Proprietary APIs

Applications based on this approach are flexible and can be fairly complex, and they also present high-level performance. Due to this flexibility, it is possible to develop the same kind of algorithms as those of the CGI approach in order to support some degree of scalability.

However, because these applications are now executed in the same address space as that of the Web server, this implies the Web server will crash whenever a crash occurs in the application. (This can be prevented by providing separate address spaces, like Oracle's Web server, but then performance and scalability cannot be the same.) Also, these applications require higher technical skills than other approaches and this implies a higher cost and risk of development, management and maintenance.

3. Client-Centric WIS

After the deployment and generalized use of the first real-world server-centric WIS, their main limitations became apparent: difficult to support complex and/or long transactions; low performance; and poor end-user interaction. This new approach, based on client-centric WIS, appeared as a solution to these limitations. (Although, as we will see, they could not solve all of them and also generated new problems.) Client-centric WIS can be divided into two distinct groups: applications based on *previously installed code*; and applications based on *mobile code*.

3.1 Previously Installed Code

In order to extend the Web client with new facilities, it is necessary to provide an API so that it can communicate with the application providing them. CCI (Common Client Interface) [NCSA95b] and CCI++ (Constituent Component Interface ++) [AM95] are examples of APIs that were proposed to provide a better cohesion and integration between the Web client and specific applications.

Based on CCI++, Netscape provides a proprietary API to enable third-party developers to write external applications that run tightly connected to its Web client (Netscape Navigator).

These applications are designated in market parlance by *plug-ins* [Net95a] and are just dynamic link executable modules (DLLs in Windows) that follow the Netscape proposed API. These plug-ins are previously installed in the client computer, and are typically responsible for handling one or more multimedia document formats (MIME types). The process works like this: whenever Netscape receives a non-standard MIME document (identified as such in the document header) the corresponding plug-in – that was previously installed and registered by the Netscape user – is dynamically loaded in the same computational context of the Web client. (This decision is responsible for some crashes if the plug-in is not robust enough.)

Figure 4 puts an emphasis (gray color) on the following components: the plug-in (an application previously installed); the MIME document(s); and some other application executed on the server machine that is responsible by delivering the MIME document (e.g., file server, video producer).

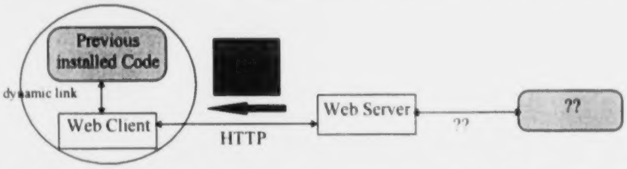


Figure 4: WIS computational model based on previous installed code approach.

Applications that follow this approach are typically very general, complex and produced by specialized software houses. Examples include viewers (e.g., for Microsoft Word documents), players (real-time sound and video) and editors of well-known multimedia formats (e.g., Acrobat's PDF, VRML, MPEG2, and so on). Also included in the category are more complex applications such as spreadsheets editors and byte-code interpreters (also called virtual machines). For example, there are plug-ins for running Java byte-codes provided by Netscape, Microsoft and Colusa.

3.2 Mobile Code

The plug-in approach already supports a reasonable integration between a Web client and its external applications. However, this approach still presents some limitations, especially regarding flexibility and portability. In addition, plug-ins require a previous installation made by the user, further complicated by the need to update for newer software versions (from alpha to beta, from beta to final, from 1.0 to 2.0, and so on).

A more general WIS approach is based on mobile code [Con95,BTV96]. This approach is based on a program stored in some computer (typically the Web server) which can be transportable (or copied, moved, etc.) across the network to another computer in order to be executed remotely.

The mobile program is written in virtually any programming language, although some languages are more suitable than others to achieve this task. For example, interpreted (scripting) languages are inherently more portable then traditional (compiled) languages because an executable is based on the machine's native architecture and will not run in a different machine architecture. However, there is an obvious loss of performance if script languages are used directly without any compilation.

Figure 5 shows the architecture of the mobile code approach. The code for the specific application (mobile program) is stored on the Web server machine (the '??' block) and is transferred on-demand to the Web client machine where it is executed – typically by a "virtual machine" block. The client has to provide a virtual machine responsible for the safe interpretation and execution of the code received. (However, if the mobile program is native code, then the it will run directly on top of the operating system.)

BEST COPY AVAILABLE

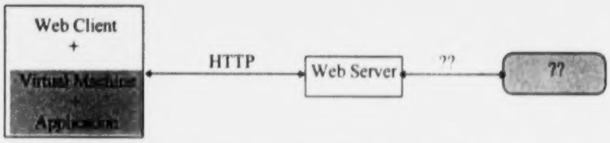


Figure 5: WIS computational model based on mobile code approach.

We found basically two distinct WIS approaches based on mobile code: mobile code embedded in HTML documents; and mobile code independent of (separated from) HTML documents.

In the first approach – *mobile code embedded in HTML documents approach* – the Web client, beyond its regular capabilities of parsing and rendering HTML elements, also needs capabilities for interpreting and executing embedded code (usually as source code). This approach has two main advantages: it allows the construction of small and simple WIS with reasonable end-user interaction, and supports the integration between HTML documents and other Web technologies (Java applets, plug-ins, etc.) what has been called "glue technology".

The majority of scripting languages in use today on the Web were either adapted to be used under this approach or originally designed for it. Examples include JavaScript [Net95b], VBScript [Mic96c], Tcl [Ous94], and Obliq [Car95].

In the second approach – *mobile code independent of HTML documents approach* – there is usually a virtual machine that executes the code referred in an HTML document but located in a separate file (usually stored on the same server machine, although not strictly necessary). As a consequence, the code can be referred several times inside the same HTML document or even in many documents. Figure 6 show how it all works.

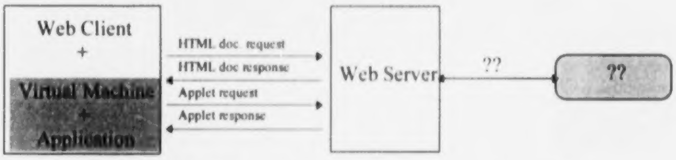


Figure 6: Mobile code, HTML document independent.

The mobile code is transferred and executed in text or binary format, and it should be noted that nothing prevents a program written in a script language to be used in a similar manner. The only apparent reason for this distinction is that script languages are represented as text and as such can be included inside an HTML document, while a general compiled program cannot.

Examples of this approach include Java applets [AG96] and Active-X controls. There are also many research prototypes based on this approach, such as Tcl [TB96] and Obliq [BN96].

3.3 Comparative Analysis

In this section we summarize and compare all client-centric WIS approaches described before. In addition to the criteria analyzed in section 2.4, security will also be analyzed here.

Analysis criteria	Previously Installed Code	Mobile Code Embedded in HTML	Mobile Code Independent of HTML
Application performance	☺	☹	😐
Application safety	☹	😐	😐
Application flexibility	☺	☹	😐
Easy development	☹	☺	☹
Safety/Security	☺	😐	😐
Application maintenance	☹	😐	☺
Vendor-neutral	☹	😐	☺

Table 2: Comparative analysis of client-centric WIS approaches.

WIS applications based on *previously installed code* present high-level performance mainly due to these reasons: the application code is already in the client machine when it is invoked and perhaps it is even loaded already in memory; and the code is compiled to the corresponding machine architecture and runs natively in the client computer. But since the application shares the memory space with the Web client, this approach also presents some safety problems, although it doesn't raise any security problem.

On the other hand, this approach requires a programmer with high-level technical knowledge to develop these specialized applications and also requires the user to manually install the program and maintain its versions. Finally, this approach is based on proprietary Web client APIs.

Mobile code embedded in HTML documents.

BEST COPY AVAILABLE

Applications based on present low to medium-level performance mainly due to two reasons. First, the code has to be transferred from the server to the client machine. Second, the code has to be interpreted directly from its text format and this is highly inefficient.

Furthermore, because these programs are downloaded from the Web server on-demand, their operations should be restricted for security reasons. (Just imagine an application that, when downloaded, removes all files in your hard disk.) For example, it is typical to restrict these applications from accessing the local file system or from establishing arbitrary network connections. However, there are still no good solutions for the resource (memory and CPU) exhaustion syndrome. Consequently, these applications should be over-restrictive and as a result are not very flexible.

On the other hand, their version maintenance and management is easily done. They did not receive the highest mark because small changes in an application may eventually require several changes in several HTML documents, but this can be solved by storing the code separately and simply referring to it from within the HTML document.

Mobile code independent of HTML documents.

Applications based on this approach also present, depending on the application and the adopted technology, a low to medium-level performance. This is because the code also has to be transferred (and, sometimes, also interpreted).

Unlike the previous model, the code usually runs natively or is interpreted in a more efficient format – virtual machine byte-code, something half-way between source code and native format. There are also many ways to improve the performance of these applications, for example, Java applets have been improved with so-called "just-in-time compilation" (dynamic compilation just before execution) and operating systems that support Java byte-code natively (and, more recently, even micro-processors designed exclusively for Java).

One of the most important strengths of this class of applications is its easy maintenance and management due to the fact that all code is stored centrally in just one place. However, the same security issues raised for the previous approach (mobile code embedded in HTML documents) also have the same solutions and problems as discussed above. Consequently, these applications present a medium level in flexibility and versatility.

Lastly, in spite of several announcements about development environments based on (Delphi-like) visual interfaces, this approach still requires a high level of technical skills. (The difficulties are raised not by the language itself, since Java is simple, but as a consequence of the many libraries that are needed for any Java program.) Lastly, due to Java success and its support by the majority of software companies, applications based on Java are in general vendor-neutral. However, the recent initiative by Sun and IBM – called "100% Java" – is perhaps a first sign that this neutrality is being put under serious challenge by Microsoft and others.

ActiveX controls

Before finishing the section, a special note should be written about the Active-X technology from Microsoft since it is so well-known. From our point of view, ActiveX can be considered an hybrid approach. From the mobility (remote transfer) and automatic configuration perspective, they are equivalent to Java applets: they are downloaded on-demand by the Web client when it founds a reference in a HTML document. On the other hand, from the portability and flexibility perspective, Active-X controls are similar to Netscape plug-ins: they are general applications that run natively and will probably be resident in the local machine (since previously downloaded controls are cached locally for efficiency reasons).

Consequently, applications based on Active-X controls present high-level performance and do not require elaborate maintenance or management. However, since controls run natively and have access to all computer resources, they present several security problems. Microsoft has partially solved this problem by digitally signing its own and other ActiveX controls. This gives some guaranties – the user knows who or which company wrote the control and can always choose not to execute it – but does not solve other security problems. (For example, these days many people are suspicious of less well-known companies that may include a Microsoft recently acquired company or a subsidiary.)

4. WIS Based on Distributed Infrastructures

The client-centric WIS approach solves some of the problems found in the server-centric approach, especially those related with performance and end-user interaction. However, it still does not give an adequate answer to the remaining issue: how to handle complex transactions between the end-user (at the client side), the application logic (now at the client side) and the database (at the server side).

This is because the HTTP protocol will never support efficient client-server communication – at least not until a new version of HTTP is widely accepted and deployed – due to its inherent connection-less mode. For the very same reason, HTTP cannot cope well with certain classes of services. These include applications involving real-time multimedia (e.g., real-time video or audio) and those accessing database systems for anything more complex than just reading a few simple records. Examples of this last class of applications are those based on long (eventually nested and/or distributed) transactions and those that manage large sets of very complex data, such as geographical or temporal information. For these kind of applications, the simple HTTP protocol and its HTML associated format are simply not enough!

As a consequence, a new approach was proposed to handle the kind of complex applications not well supported by the previous approaches described above. In this new approach, depicted in Figure 7, the Web client and the Web server are just used for the initial interaction, i.e. to establish the connection and/or to transfer the corresponding (mobile code) application. After that, the client (non-Web) application establishes another (non-HTTP) connection with its proprietary (non-Web) server. This means the application itself is not actually based on the Web since it does not need the Web infra-structure anymore.

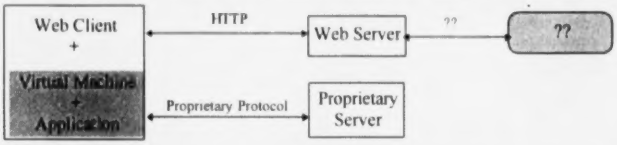


Figure 7: WIS computational model based on combined approaches.

The main advantage of this approach is obvious: once free from the standard Web shortcomings, the application and its server can now take full advantage of well-known distributed systems technology like message passing, RPC and other proprietary communication protocols. However, this approach creates difficulties regarding the integration and interoperability between different applications that follow this approach. In particular, each application will be incompatible with any other application!
As a result, a number of compromises were found based on current Web technology and a (more or less) standard distributed infrastructure. Figure 8 shows the basic idea in which the Web is used as a worldwide mechanism for finding and accessing a given application. In addition, the Web client is still used to provide a consistent human-machine interaction based on HTML and/or Java windowing toolkit [Yu96].
In this distributed approach, the application is effectively divided between a client and a server sub-application. The only difference to a standard Web application is that applications based on this new approach now communicate via the distributed infrastructure (instead of using HTTP and HTML) for efficiency reasons. Below we briefly describe the two main technologies that are used in these environments, respectively based on CORBA and Java.

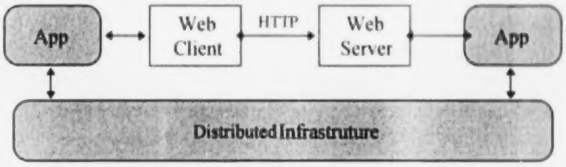


Figure 8: WIS computational model supported by a common infrastructure.

4.1 Based on CORBA

The *Common Object Request Broker Architecture* (CORBA) [Spi96] seems to be a likely solution to the approach depicted in Figure 8. CORBA is a standard model for what an RPC-like communication mechanism should offer at the application level, and has been implemented by an (increasing) number of software vendors. The latest CORBA 2.0 specification even proposes a standard communication between different CORBA implementations so that all CORBA products will eventually inter-operate between them.
A solution based on CORBA and the Web effectively integrates the two most popular distributed technologies in use today and seems a excellent candidate for developing distributed applications. Figure 9 depicts the CORBA approach to Web development, in which Java can be used at the client side as an add-on to the Web.

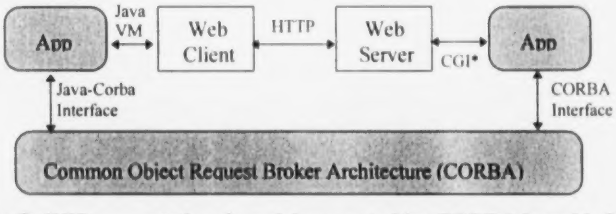


Figure 9: WIS computational model supported by CORBA based infrastructure.

It should be noted here that CORBA needs some programming capability at the client side – so Java or some other programming language will have to be used – but that is only natural if we assume this approach was proposed precisely because some programming was needed at the client! As an alternative,

BEST COPY AVAILABLE

JavaScript or VBscript could be used if the Web client itself supports CORBA, as Netscape said will happen with the next version of Navigator (Microsoft will propose something similar for sure).

So let us assume the application uses Java at the client side. The client application is a Java applet being executed in the context of the Web client virtual machine. In order to access its server-side application via CORBA, the applet either includes CORBA itself or relies on CORBA support from the Web browser as we discussed in the previous paragraph. At the time this paper was written, there was no commercially available Web browser supporting CORBA. There are, however, many Java-based CORBA products such as JavaSoft's JOE, PostModern Computing's Black Widow, or Iona's OrbixWeb.

At the server side, the application provides a set of services accessible via its CORBA interface. There are many CORBA-compliant products available commercially today, so in principle any one of these products could be used. However, if the application is designed to support services to client applications using other CORBA products, than it should be based on the standard IIOP protocol for interoperability between different CORBA implementations.

4.2 Based on Java

CORBA was designed to support communication between applications by taking advantage of the best technology that existed in the early 90s. It has been widely implemented and used in a large number of industrial, mission-critical distributed applications. As a result, CORBA is praised by its virtues in the distributed systems community, with entire conferences on the latest CORBA products. However, there is an interesting argument posed by the Java community regarding the CORBA approach, as follows.

- Java represents a newer (read better) technology and was designed, like CORBA, for building distributed applications, so why use CORBA if we now have Java?
- In addition, Java is already supported by most Web clients and it has built-in support for inter-process communication, so why support CORBA in addition to Java?

The CORBA people argue that CORBA is a proved technology. According to them, Java can be an interesting programming language (like Tcl, Smalltalk, or Obliq) but "real programs" are written in "real languages" (like C or perhaps C++). Also, there is nothing in Java like the world-wide distributed applications built with CORBA. Furthermore, the CORBA interface is (at least, in theory) language independent and so CORBA can be used by any programming language, including Java. For example, an existing server application written in C can be used by a Web client application written in Java provided both use CORBA.

Here we are interested on approaches for building Web applications from a technological point of view. In this paper it does not matter if Java or CORBA will eventually succeed (or maybe both) since the issue here is how to build Web applications based on a distributed infrastructure using Java and compare the result with CORBA.

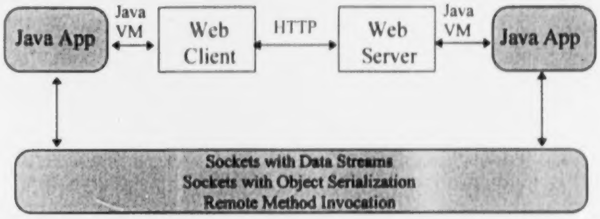


Figure 10: WIS computational model supported by Java based infrastructure.

Figure 10 depicts the distributed infrastructure approach based on Java. At the client side, the application consists of a Java applet that is executed by the Web client virtual machine. Beyond its basic functionality such as end-user interaction, it can invoke services provided by another Java application at the server side. The communication between the client and the server applications is supported and managed by a distributed infrastructure based on Java.

The most popular approaches for communication between two Java processes is by means of raw sockets or the Remote Method Invocation (RMI) protocol [Jav96a]. RMI is a proprietary (just for Java) CORBA-like RPC mechanism for calling methods on remote objects. The arguments are passed by copy using Object Serialization, a mechanism that can be used independently, e.g. for storing objects on files or copying them via sockets to another process.

If the Web server is itself written in Java – such as the Java Web Server from SunSoft and many others – then programmers have enormous opportunities for taking advantage of increased efficiency, integration and cooperation between the Web server and the sever-side application. For example, SunSoft is promoting this integration with its idea of servlets, Java programs that can be used to extend the basic Web server functionality. Further integration can be achieved if the Web client is also written in Java, such as HotJava from SunSoft.

BEST COPY AVAILABLE

5. Summary

In this paper we presented the basic technological approaches for designing and building Web information systems. Two complementary approaches were identified: WIS whose activity is mainly executed at the server side (Server-centric WIS); and WIS whose activity is mainly executed at the client side (Client-centric WIS). Furthermore, we identified a third approach, that integrates the previous two, based on a distributed infrastructure.

All these approaches are based on existing technology, sometimes widely used in other contexts, and as a result there has been an enormous development activity all over the world building Web information systems. There are already a number of mission-critical WIS, especially in restricted contexts – often inside an organization or a dispersed but well defined community. Examples include: collaborative work intra- and inter-organizations; public information kiosks; support for the sales people; and document storage and retrieval, including digital libraries.

However, some limitations are now becoming visible. For example, server-centric WIS typically offer a *poor interaction with the end-user* based on forms; *have difficulties supporting complex transactions* like reliable debit-credit operations; and *present a low-level performance* due to its centralized architecture and an inefficient HTTP protocol.

Client-centric WIS were developed to eliminate these problems, but also raised new issues regarding downloading time, version maintenance, security and safety.

Although the new distributed approach already permits to build WIS applications with a quality similar to traditional client/server applications, it raises new problems regarding the large number of proprietary products being sold by many vendors in a highly competitive market.

Some of the questions that are now being posed include: Is it possible to build high-quality WIS using only standard technologies and products like HTML and Java? If not, is it possible to achieve a reasonable level of interoperability amongst WIS build with different technologies? We hope that Java is the answer to all these questions, but surely only the market knows what will happen.

References

[AG96] K. Arnold, J. Gosling. The Java Series - The Java Programming Language. Addison-Wesley Publishing, 1996.

[AM95] C. Ang, D. Martin. *Constituent Component Interface ++*. Centre for Knowledge Management, University of California, 1995.
`http://www.eolas.com/eolas/ccippapi.ht m`

[BN96] M. Brown, M. Najork. Distributed Active Objects. Fifth International WWW Conference, Paris, *Computer Networks and ISDN Systems*. 28 (7-11), 1996.

[BTV96] J. Baumann, C. Tschudin, J. Vitek, editors. *Proceedings of the 2^{nd} ECOOP Workshop on Mobile Object Systems (Linz, Austria)*. Dpunkt, 1996.

[Car95] L. Cardelli. Obliq: a Language with distributed scope. *Digital White Paper*, Digital Equipment Corporation, Systems Research Center, 1995.
`http://www.research.digital.com/SRC/Obliq/Obliq.ps`

[Con95] D. Connoly. *Mobile Code - Issues on the Development of Hypermedia Applications*. W3C, 1995.
`http://www.w3.org/pub/WWW/MobileCode`

[DSCOB95] J. Delgado, N. Santos, P. Caetano, L. Osório, J. Barros. Extensão das capacidades de um cliente WWW com um sistema telemático multimédia. In *Proceedings of the 1st WWW Portuguese Conference*, Braga, Portugal, 1995.

[Jav96a] JavaSoft, a Sun Microsystems Business. *Java IDL*. 1996
`http://splash.javasoft.com/JavaIDL-alpha2.0/`

[Jav96b] JavaSoft, a Sun Microsystems Business. *Java Beans: A Component Architecture for Java*. 1996
`http://splash.javasoft.com/beans/WhitePaper.html`

[LSW95] S. Lucco, O. Sharp, R. Wahbe. Omniware: a Universal Substrate for Web Programming. Fourth International WWW Conference, Boston, In *WWW Journal*. W3C, Dec. 1995.
`http://www.w3.org/pub/Conferences/WWW4/Papers/165/`

[Mic96a] Microsoft Corporation. *IDC - Internet Database Connector*.
`http://www.microsoft.com/`

[Mic96b] Microsoft Corporation. *What is Active-X?* July 1996.
`http://www.microsoft.com/ activex/actx-gen/awhatis.htm`

[Mic96c] Microsoft Corporation. *Visual Basic Scripting Edition*. 1996.
`http://www.microsoft.com/vbscript/vbsmain.htm`

[NCSA95a] NCSA, University of Illinois. *Server Side Includes (SSI)*. 1995.
`http://hoohoo.ncsa.uiuc.edu/docs/tutorials/includes.html`

[NCSA95b] NCSA, University of Illinois. *Common Client Interface Protocol Specification*. 1995.
`http://yahoo.ncsa.uiuc.edu/mosaic/cci.spec.html`

[MdS96] M. Mira da Silva. *Models of Higher-order, Type-safe, Distributed Computation over Autonomous Persistent Object Stores*. PhD Thesis, University of Glasgow, 1996.

[MdSS97] M. Mira da Silva, A. Silva. Insisting on Persistent Mobile Agent Systems with an Example Application Area. In *Proceedings of the Mobile Agent Workshop, Berlin, Germany, 1997*.

[Net95a] Netscape Communications. *Netscape Plug-ins*. 1995.
`http://home.netscape.com/eng/mozzila/2.0/handbook/plugins`

[Net95b] Netscape Communications. *JavaScript Authoring Guide*. 1995.
`http://home.netscape.com/comprod/products/navigator/version _2.0/script /script-info/index.html`

[Ous94] J. Ousterhout. *Tcl and Tk Toolkit*. Addison-Wesley Publishing, 1994.

[Ques96] Questar Microsystems Inc. *Advanced HTTP Server Side Functionality with SSI+*. 1996.

[Rob96] D. Robinson. *The WWW Common Gateway Interface Version 1.1*. Internet Draft, 1996.
`http://www.ast.cam.ac.uk/~drtr/cgi-spec.html`

[SAD96] A. Silva, G. Andrade, J. Delgado. A multimedia database supporting a generic computer based quality management system. In *Proceedings of the 9th ERCIM Database Research Group Workshop*, Darmstadt, Germany, 1996.
`http://bruxelas.inesc.pt/~alb/papers/edrg95.ps.gz`

[SBD95] A. Silva, J. Borbinha, J. Delgado. Organizational management system in an heterogeneous environment - A WWW case study. In *Proceedings of the IFIP working conference on information systems development for decentralized organizations*, Trondheim, Norway, 1995.
`http://bruxelas.inesc.pt/~alb/papers/ofw3.ps.gz`

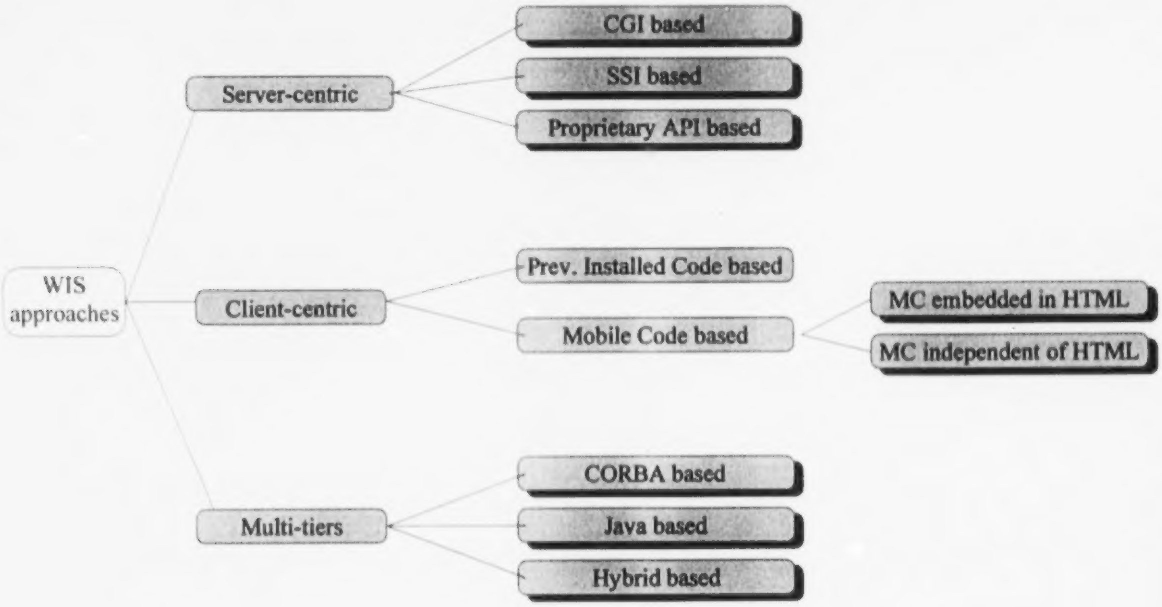
[Sie96] J. Spiegel. CORBA Fundamentals and Programming. Wiley, 1996.
See also `http://www.omg.org/corfun/corfunhp.htm`

[SMdSD97] A. Silva, M. Mira da Silva, J. Delgado. AgentSpace: Motivation and Requirements for the AgentSpace: A Framework for Developing Agent Programming Systems. *Proceedings of the Fourth International Conference on Intelligence in Services and Networks* (IS&N'97), Cernobbio, Italy, 1997.

[TB96] P. Thistlewaite, S. Ball. Active Forms. Fifth International WWW Conference, Paris, *Computer Networks and ISDN Systems*, 28 (7-11), 1996.

[Yu96] N. Yu. *AWT Tutorial*. University of Alberta, 1996.
`http://ugweb.cs.ualberta.ca/~nelson/javaAWT.Tutorial.html`

Annex: Schematic hierarchy of the identified WIS approaches.


WIS approaches
Server-centric
CGI based
SSI based
Proprietary API based
Client-centric
Prev. Installed Code based
Mobile Code based
MC embedded in HTML
MC independent of HTML
Multi-tiers
CORBA based
Java based
Hybrid based

BEST COPY AVAILABLE

Design Considerations in Converting a Standup Training Class to Web-based Training: *Some Guidelines from Cognitive Flexibility Theory*

Nancee Simonson

Training Specialist
The Bureau of National Affairs, Inc.
Washington, D.C. U.S.A.
nsimonson@bna.com

Abstract

This project was to convert a standup training class to an interactive web site to be posted on our Intranet for access by all employees. The class objective is to familiarize all employees with the three-branch federal government processes and the associated documents that go with the steps in thegovernment activities. The reasons for Web-delivery are the wide range of possible users, the difficulty of delivering stand up training to such a wide audience with different locations and time schedules, the "just in time" need to know data of this type for completing job tasks, and the availability of much of the source material now on-line. Cognitive Flexibility Theory leads to an instruction theory which calls for flexible learning environments that require multiple representations of items, repetition, active learner involvement, avoidance of oversimplification of content, situated cases, and multiple interconnections. This type of learning/instruction is particularly enhanced/facilitated by the capabilities of computer technology delivery and, in particular, hypertext.

Theory

Cognitive flexibility theory focuses on the nature of learning in complex and ill-structured domains. Spiro & Jehng [Spiro & Jehng 1990] p. 165, state: "By cognitive flexibility, we mean the ability to spontaneously restructure one's knowledge, in many ways, in adaptive response to radically changing situational demands...This is a function of both the way knowledge is represented (e.g., along multiple rather single conceptual dimensions) and the processes that operate on those mental representations (e.g., processes of schema assembly rather than intact schema retrieval)."

Cognitive flexibility theory is based on the constructivist theory of learning. It is premised on the idea that many--if not most--knowledge domains are complex and ill-structured. It attempts to solve known problems of learning failure and learning transfer. [Spiro et. al. 1991]

Jacobson outlines the features of cognitive flexibility theory: complex knowledge may be better learned for flexible application in new contexts by employing case-based learning environments that include features such as: (a) use of multiple knowledge representations, (b) link abstract concepts in cases to depict knowledge-in-use, (c) demonstrate the conceptual interconnectedness or web-like nature of complex knowledge, (d) emphasize knowledge assembly rather than reproductive memory, (e) introduce both conceptual complexity and domain complexity early, and (f) promote active student learning. [Jacobson et. al. 1996]

Spiro [Spiro et. al. 1991] outlines a number of features of the theory: 1. the importance of multiple positions of instructional content, from multiple organizational schemas for presenting subject matter to multiple representations of knowledge, 2. importance of students' active participation, 3. revisiting the same material at different times in rearranged context for different purposes and from different conceptual perspectives, 4. avoid oversimplification.

Further, the theory suggests a criss-crossed landscape via a nonlinear and multidimensional traversal of the complex subject matter [Spiro et. al. 1991] The Internet and hypertext/hypermedia are particularly suitable for applying

© Nancee Simonson, 1997

cognitive flexibility features because the media easily supports presentation in multiple perspectives and knowledge criss-crossing . [Spiro et. al. 1991] , [McManus 1996]

Technology environment such as the Internet can present the learning with the world in its natural complexity; rather than simplifying situations or tasks, an on-line environment can embed the learning in the real-world situations, (Jonassen) "We believe that hypertext is among the best examples of contructivistic learning environments, because acquiring knowledge from hypertext requires the user to engage in constructivistic learning processes. Learning from hypertext is task driven. It depends largely on the purpose for using the hypertext, which in turn drives the level of processing." [Jonassen] He says that cognitive flexibility theory is and effective way for accomplishing the goals because it is case-based and involves meaningful world tasks.

McManus [McManus 1996] developed a Hypermedia Design Model based on Cognitive Flexibility Theory. The steps of this design model different from tradition instructional design : 1. Define the learning domain, 2. Identify cases within the domain, 3. Identify themes/perspectives to be highlighted, 4. Map multiple paths through cases to show themes, 5. Provide learner controlled access to cases, 6. encourage learner self-reflection.

Rationale: Cognitive Flexibility Theory as proposed by Spiro, Feltovitch, Jacobson and Coulson [Spiro et. al. 1991] is a constructivist theory which argues that the complexity of real world situations, the "ill-structuredness" of most knowledge domains, and the failure of transferring learning in traditional ways poses serious problems for traditional theories of learning and instruction. While the traditional theories may be applied to novice learning, almost opposite techniques are required for learners to develop the cognitive flexibility to apply learning to complex, unrelated real life situations in more advanced learning.

In order to achieve this cognitive flexibility, the learning theory leads to an instruction theory which calls for flexible learning environments that require multiple representations of items, repetition, active learner involvement, avoidance of oversimplification of content, situated cases, and multiple interconnections. "We have called the instructional theory that is derived from Cognitive Flexibility Theory and applied inflexible computer learning environments Random Access Instruction [Spiro et. al. 1991]. This type of learning/instruction is particularly enhanced/facilitated by the capabilities of computer technology delivery and, in particular, hypertext.

My product deals with an "ill-structured domain," i.e. the federal government process. While, in its simplified form, it might appear that knowing the steps of the government process is pretty linear and straightforward and not ill-structured at all. However, Cognitive Flexibility Theory says precisely that over-simplification leads to lack of transference. Even though in context this product is a "basic" class, the fact is that learners must use the information for complex job tasks--such as answering customers unpredictable questions for information of all sorts, analyzing and writing about current developments in government or researching specific cases. In this setting, then, the traditional high school government or civics course could be considered the "novice" level learning, in which the structure is simplified. But in this "basic" class the principles of ill-structuredness and complexity apply.

Because the theory is concerned with addressing failures of learning transfer through oversimplification and over-structuring, I am particularly interested in its application to our training for employees to understand the federal government process. While the principle of learning transfer may not apply perse-- Our employees don't need to memorize the steps in the government process--they must have a degree of knowledge of the structure of the domain to be able to quickly go to specific information they need quickly. They need to know how to apply the information in specific job situations that are very diverse: A reporter needs to know the underlying process in order to analyze a current event, devise probing reporting questions, gather the proper data, talk to the right sources. A research person needs quick retrieval abilities of specific data when faced with a customers questions or access to phone numbers or organizational charts to determine where to get the information. A clerical may need to have a mental model of the structure for filing data related to publications content. A sales person may need only a quick outline of information to field customer questions and understand basic publication content for a wide variety of publications. So, the goal is NOT to be able to recite the steps of the process, rather use it as a tool to analyze and gather appropriate additional data and apply it to judgments about the worth and usefulness and likely next steps of that data for our subscribers.

The case study approach applies to the way users use the knowledge, since on-the-job use is always a different "case" pertaining to the federal government process. I did not use it as the instructional model, however, since it would simply take too long. I found with our users since they have been exposed to the job tasks for some time have constructed their own models of how to organize useful information about the process--whether it be by process steps, by resource (phone numbers, addresses, directories, etc.) or by institutional organization. Therefore the goal was to

© Nancee Simonson, 1997

replicate this work process as much as possible to make information retrieval as quick as possible. In this respect the domain based on usage, and not strictly knowledge, is complex and ill-structured as well.

Project

The project was to convert a standup training class to an interactive web site to be posted on our Intranet for access by all employees. The class objective is to familiarize all employees with the three-branch federal government processes and the associated documents that go with the steps in the government activities. The reason is that this is the basis of our corporate product and activities: we are a private publisher that gathers information on government activities and developments for our professional customers. So most of our employees need to know something about federal government processes, though each in different ways.

The reason for putting something like this on the web is the wide range of possible users, the difficulty of delivering stand up training to such a wide audience with different locations and time schedules, the "just in time" need to know data of this type for completing job tasks, and the availability of much of the source material now on-line. The project was originally conceived as Computer-based training, with tutorials and quizzes, but after further study and needs assessment, I realized that it would be more useful as an electronic performance support system and, indeed, has been extremely well-received in that respect!

The project was done in Word with its add-on Internet Assistant. This was "within budget" because Word is our office standard and Internet Assistant is free. It is an adequate software package for this project, especially since it lets you code by hand if necessary. The product is posted on our Internal Home Page, with a direct link from the first page for ease of use. Most of the intended users have access to the Internal Home Page (i.e. hardware and software). For those that don't it is available on computers in the centralized Individual Learning Center, or a copy is available on a disk for someone to take home or use at another computer of their choice. Over the next five years, it is anticipated that all those in the intended audience will have desktop access to the Internal Home Page.

Design

User Interface--We originally planned a frames-based interface and a "site map" of the main sections and subsections in order to keep users in the product and provide an easy way to navigate. However, use of this is being suspended until this more advanced feature is needed (see discussion in changes section below) The program is divided into three sections: congressional (legislative), executive (regulatory) and judicial (courts.) Plus a brief, entertaining and optional introduction, fully linked index and resource guide. Colors are limited to white (background), blue (highlights and graphics) and green (optional). The choice of colors is for their meanings of tranquillity and steadiness. The navigational buttons will be plain blue. The graphics are basically black and white drawings, with some blue highlights, of three federal buildings: Capitol, White House and Supreme Court to symbolize the three branches of government. The main frame consists of a horizontal bar with 6 buttons for the intro, three branches, index and resources. Also two site maps for the two themes of the program: the sequential government process and the arrangement of corresponding documents will be added later. The six segments will be contained in six scrollable documents: each section will be in one file, rather than using a stacking card arrangement with one screen per page.

Content outline:

Intro: a brief introduction of the course with link to Schoolhouse Rock video WWW site on-line: "Preamble" and "Three Ring Government." These video segments were used in the class, but the on-line site contains audio and lyrics and some video. It's entertaining, a fun "mini-case" but is separate optional button since it is not vital to everyone.

Legislative: steps of the process with links to supporting documents, flow charts at each step of the way.

Executive: White House and Regulatory steps of the process from bill signing to implement regulations with links to supporting documents and flow charts at each step of the way.

Judicial: Supreme court and federal court process step-by-step with links to supporting documents and flow chart at each step of the way.

Resources: list of resources on line.

© Nancee Simonson, 1997

Index: basically a map of the program in index form linked to the segments of the program so a user could go immediately to a section or document needed at a specific time.

Improved design:

One benefit of a Web-based delivery mechanism for this course is the ability to anchor the instruction to links to current and live documents in the federal government process that are on-line. Also to link to host sites that explain themselves (such as Congress, the courts and the White House) rather than have an uninvolved third party, no matter how knowledgeable, provide the content.

But even further, Cognitive Flexibility Theory basically adds theory and design strategy to the technology driven delivery mechanism of hypertext or CBT on the Web. The theoretical elements determine best designs for the Web-based instruction.

According to the theory, flexible learning environments permit the same items of knowledge to be presented and learned in a variety of different ways and for a variety of different purposes. For my project, the material is presented as an overview from the perspective of steps in a process. At the same time, the hypertext capabilities with the Cognitive Flexibility Theory applied, creates a richly cross-linked database that delivers a series of sequential documents, discrete moments in the process, a list of resources, relationships among the processes, and deviations from the structured steps.

This is useful, because the audience really is the whole company, since everyone in the company needs to be familiar with the federal processes. However, discrete groups will use it for different purposes. Our research department answer customer questions, so they will need to be able to look up discrete events or documents. A new hire may want to study the process sequentially. A reporter may need parts of the processes depending on a given assignment at a given time. So that the navigational design must provide access from all these perspectives, and navigation along all these lines.

Thus, a concept map or site map in frames will be used eventually that shows possible navigation paths from several different viewpoints: linear overview, process content and documents. The index itself will also provide a linear overview.

In addition, more links were added to give access to documents in different ways. For instance, a link was included to a lengthy, detailed document on the judicial process for those who wanted to "study" it further. However, it turned out, others wanted immediate access to some of the more obscure courts without wading through all of the text and the entire process. So links were added where the summary on the first page mentioned them. No additional information was given on the first page, because it was still intended to be a quick overview. Another addition was direct links to a directly of courts and phone numbers for those who would use it as a job aid in doing their research. Again, it was the use of links in a variety of ways that permitted many cross paths for navigation. We needed to provide depth at the same time that we needed to provide immediate access to some details for different types of users. The use of hypertext and well-designed multi-navigational paths proved a highly effective way to do this and, at the same time keep the design simple and elegant.

Another advantage, applying this theory is linking the material to "live" documents and "live" sites--such as the White House, National Archive, House and Senate. This is not available easily in a standup training class where you would prepare handout material in advance and once it's on paper it is "dead."

The hypertext media also allows for more extensive use of case studies--in this product, that would be actual legislative or judicial histories--through multiple links, without overwhelming the user or detracting from the simplicity of the design. It also leaves the use of the cases in the control of the user.

The only place I would differ in conclusions with the theory is that the authors argue that it is best applied for "advanced" learning. Essentially, in my case, I used principles of the theory to turn a "basic" class into a dual purpose product: that is, the framework or basic interface continues to be an overview of the government processes. But robust hypertext design permits inclusion of very advanced material and job aids without detracting from the basic outline.

Implementation and Testing

Implementation description. The implementation, simply, was posting it on our Internal Home Page. The Internal Home Page contains a link to "Editorial Training".

© Nancee Simonson, 1997

In order to make it easy to access , right under the Editorial Training heading are three links: 1. list of classes. 2. registration form, and 3. Introduction to Government Process Class. I wanted people to have immediate awareness of the product and not have to drill down to get it. While the class is also linked to the second level where it is described in the schedule, it had to be on the front page as well because it is and EPSS and not simply class materials.

Since in the time allotted for this class, we only completed the intro page and one of the three sections--Judicial Process--the other two sections were "under construction."

Promotion consisted of : Announcement to Executive Editor at a meeting, then second ranking editorial executives, then a general announcement to all employees.

At the same time, testing by users continues: this provides continuous improvement, and, at the same time, a marketing tool. For example, the legal trainer for the research department is already using the Judicial site for her training before the "ink" was barely dry on the coding to get it on the internal home page! I find that the ability to use this product "in pieces", so to speak,---rather than using the whole package as you do with a stand up class--to be one of the major benefits of having this as a Web product or EPSS. Users really can take what they want and use it however it best suits their purposes.

Testing and evaluation results. Testing was done continuously by prospective users of the product throughout the company: research, new hires, nonlegal editorial, legal editorial, content experts, managers of prospective users, executive editor. We tested for a variety of situations and a variety of users. 1. We tested the product with the training staff for its instructional design effect and usability with completely new users. 2. We tested immediately after a stand-up training class so that we could specifically get comparative feedback. 3. We tested with managers of those who went to the stand-up training class to get feedback on its usefulness for their staff development purposes and also comparatively if they preferred the CBT to stand-up training for their purposes. 4. We tested with technical interface and instructional design staff. 5. We tested with the subject matter experts for content and design. 5. We tested with managers and staff randomly for usability and content. Testers were given a disk with the prototype, plus a paper printout on which they were instructed to make comments line by line on the elements of the product. For many, I also aurally received comments and discussed solutions to comments, recommendations, additions and deletions. They also had a brief one-page questionnaire with three questions about the overall usefulness, ease of use, and suggested changes.

They uniformly found the product useful. My favorite review said it was "awesome." Partly this is because it is the first "just'-in-time" training product or EPSS in the company and it is on a topic that nearly everyone needs to know something.

At the same time, testers gave A LOT of really good specific suggestions for additions and changes. Some of the critical comments concerned readability--the text in a chart was not clear--and a few typos. These were all corrected. Most of the comments concerned ADDITIONS. This indicated that people found it useful and wanted even more capabilities in the package.

Suggested changes based upon results . The suggested changes are too numerous to list all of them here.

A lot of the changes had to do with design issues: adding more information and detail without taking away from the simplicity; giving people the content they needed, where they needed it, without taking away from the purpose of the product which was to be an overview of the government process. This was solved by adding a few more sections, but mostly by arranging links to solve the access issues. For instance, one person wanted immediate access to information on more obscure tax and bankruptcy courts. While this was in a link to a more detailed discussion later, I added links to these specific sections in the list of courts.

Another person wanted to know about the Bill of Rights. It was a big question. While that did not strictly fit in with the structure--which is the three branch process--it seemed a big enough issue to this user that it was detract from the product. I made a content connection by adding a short paragraph, with a link to the text, and the explanation that this is the source of a lot of litigation, regulation and legislation in the three branches. My operating principle here is that since this will be a self-paced instructional piece , used without the guidance or control of an instructor, all individual questions are important, because they are precisely what will come up when a user uses the product on a solitary basis. As a result I tried to address EVERY question or comment by the uses in some way.

© Nancee Simonson, 1997

Others found the directories and telephone numbers in one of the lengthy documents most useful, so a link will be added to the first page of the Judicial Section that will give users up-front access without changing the overall format of the first page.

There were no real suggested changes about overall format. Again, I attribute to my goal of strictly keeping it simple, and to the newness of such a product in our workplace. I suspect as people become more familiar with this product and other Web-based products, that more suggestions will come up and more changes will be made. The intent is to continually review and update the product to keep up with technology and user needs.

The frames design was not used at this stage, since it was not needed to make use of the product easy. However, I am preparing the coding for that because I feel that as the technology advances and users get used to more sophisticated designs that will be necessary. For example, for the Internal Home Page as a whole, the design is very plain, and the designers and now only beginning to design a new tables format. This product by itself is the most animated and sophisticated item on the Internal Home Page. For consistency and ease of use, the simpler design at this point was more compatible on the Internal Home Page.

Discussion

Quality of software solution. I think the product more than met its goal of providing just in time training on government process. As mentioned, the project took on the role of an electronic performance support system. Without knowing the technical terms and concepts, using began asking for additional items to make it such. People were thrilled with the EASE OF ACCESS to the material. This was one of the goals and the response indicated it was well met. The "teacher" or subject matter expert involved is a librarian and also one of the maintainers of the Internal Home Page so that it will be easily maintained and updated.

Overall, the original stand-up training was much-needed and well-received. This product ended up going quite a bit beyond the resource provided in the stand-up training and gave people an easy desk-reference, telephone director, subject matter, content review all in one.

It can also be the beginning of building an internal "knowledge base" of frequently asked questions about the process. In other words, it is turning in a dynamic learning tool, resource and just-in-time training.

For me, it was an experience in learning the richness of the Web-format and HTML in providing a product that can answer many needs at once. Often in stand-up training have a group with diverse needs at one time really degrades the training.

Since the Web and HTML are so robust, the design can be made with links to a lot of information without corrupting the overall simplicity of the original product. It can also be made graphically attractive to use

As for project and team work--I think the loosely constructed team of experts worked very well. Each person had experience in working on teams so there was not a long learning curve in that area. We did not have lengthy meetings. We were able to discuss a lot via interoffice email and get things done efficiently . (Having teamwork experience and enthusiasm helps a whole lot on a project like this. I am not a fan of lengthy meetings and discussions.) In addition we were able to have additional resources--frames programmer, interface designer--who were not designated team members, at our disposal easily and enthusiastically! As far as lessons learned, I would do it again that way: a small dedicated team of those who will do the bulk of the work, with easy access to call on other expert resources as needed.

Conclusions

1. KEEP IT SIMPLE. One of my goals , of course, was to keep the design and presentation simple. One reason was because the Internet and Intranet are very new here and people are not that sophisticated about its capabilities. But the simplicity turned out to be the magic pill. Everyone who tested the product really thought it was valuable, learned something , found it easy to use.

2. At the same time, keeping it simple yet attractive and easy to use is VERY DIFFICULT AND A LOT OF WORK. I found the design and development implementation took a lot longer to do--to get right!--than expected. Things change on the Internet, little codes don't work, the design doesn't look right in all browsers. Which leads me to another point:

© Nancee Simonson, 1997

3. YOU MUST DESIGN FOR ALL BROWSERS! It looks different. What looks good--or a difficult execution that looks "ok" on one browser may come out entirely different on another. Again a simple design makes this easier to do, but then it's difficult to accomplish the goals in a simple design--a circular problem.

4. Above all else, YOU MUST KEEP THE USERS NEEDS IN MIND FIRST.

5. TEST and TEST and TEST SOME MORE. Especially since this is a product on the Intranet for ALL employees, I couldn't get away with overlooking things that might for a very focused group. The testing on different types of employees was so valuable because it gave insight into how each type would use it; what their particular information needs are, what questions popped up for each. For example, the editors were particularly aware of typos--might seem a small thing, but that's what slowed them down! One person in reading about the US Constitution asked about the Bill of Rights. While this was not directly related to one of the three branches of government, it was the first question that popped into his mind. So a added a paragraph and link, with the rationale that the rights in the Bill of Rights prompt a lot of the legislation, regulation, and litigation that churns through the three branches. It became a richer product for everyone's input.

6. FIND SOME WAY TO INCLUDE ALL OF THE SUGGESTED CHANGES. I discovered that precisely because this is an on-line product--not standup training--users won't have a trainer to ask answer questions. Therefore, any question they have while using the product will slow them down or turn them off if they are not answered. The testing process needs to be very thorough to uncover all of these questions and then the solutions must be incorporated into the product.

7. IT IS AN ITERATIVE PROCESS. At this point I don't think the product will ever be completely done. That's OK. I think it is extremely important to add to it what people need to do their jobs. At the same time, design issues and goals MUST be kept in mind so that the additions and changes don't change the integrity of the product.

8. Experience with TEAMWORK helps. Learning teamwork should not be a part of the project!

References

[Agada] Agada, John, "Analysis of Information Repackaging (IR_ Processes using the Instructional Systems Design (ISD) Model, School of Library and Information Science, University of Wisconsin-Milwaukee, Online: http://www.usq.edu.au/electpub/e-jist/agada.htm

[Carleton] Carleton University, The Tutorial Gateway. Online: http://www.civeng.carleton.ca/~nholtz/tut/doc/doc.html

[Eklund] Eklund, John, "Cognitive models for structuring hypermedia and implications for learning from the world-wide web." Online: http://www.scu.edu.au/ausweb95/papers/hypertext/edklund/index.html.

[Jacobson et. al. 1996] Jacobson, Michael, Maouri, Chrystalla, Mishra, Punyashloke, Kolar, Christopher. (1996). "Learning with Hypertext Learning Environments: Theory, Design, and Research," Journal of Educational Multimedia and Hypermedia, 5(3/4), 239-281. Online: http://lpsl.coe.uga.edu/jacobson/papers/JEMH96/JEMH96.html

[Jonassen] Jonassen, David, and McAleese, Terry Mayes Ray, "A Manifesto for a Constructivist Approach to Technology in Higher Education." Online, http://www.icbl.hw.ac.uk/ctl/msc/ceejw1/paper11.html

[Leigh] Leigh, Douglas, "The Internet and Distributed Learning: Instructional Designer's Medium and Tool." Online: http://mailer.fsu.edu/~dleigh/superfulx/words/dleight/NetDistLearn.html.

[McKendree 1995] McKendree, Jean, Reader, Will, and Hammond, Nick. (1995). "The "homeopathic fallacy" in learning from hypertext," in Interactions, ACM Press, Vol. 2, No. 3. pp. 74-82, 1995. On-line: http://www.ioe.ac.uk/tescwwr/Homeopathy.html

[McManus 1996] McManus, Thomas Fox. (1996) "Delivering Instruction on the World Wide Web," University of Texas at Austin. Online: http://www.edb.utexas.edu/coe/depts/ci/it/projects/wbi/wbi.html

[Mayes] Mayes, Terry, "Another Look at Courseware." Online: http://www.icbl.hw.ac.uk/ctl/mayes/paper13.html.

© Nancee Simonson, 1997

[Spiro et. al. 1991] Spiro, Rand J., Feltovich, Paul J. , Jacobson, Michael L., and Coulson, Richard L. (1991) "Cognitive Flexibility, Constructivism, and Hypertext: Random Access Instruction for Advanced Knowledge Acquisition in Ill-Structured Domains," in T. Duffy & D. Jonassen (eds.), Constructivism and the Technology of Instruction, (pp. 57-75) Hillsdale, NJ, Lawrence Erlbaum Associates. Online: http://www.ilt.columbia.edu/ilt/papers/Spiro.html

[Spiro & Jehng] Spiro, Rand J., and Jehng, Jihn-Chang, "Cognitive Flexibility and Hypertext: Theory and Technology for the Nonlinear and Multidimensional Traversal of Complex Subject Matter." (1990), in Nix, Don, and Spiro, Rand, Cognition, Education, Multimedia, Hillsdale, N.J., Lawrence Erlbaum Associates.

[SuperCal] SuperCal, Staffordshire University, A joint project of University of Trieste, Italy, and Staffordshire University, UK, on Computer Aided Learning materials.

[Institute for Learing Technologies] WWW Constructivist Project Design Guide, Institute for Learning Technologies, Teachers College Columbia University. Online: http://www.ilt.columbia.edu/k12/livetext/webcurr.html

[University of Southern Queensland] University of Southern Queensland, Instructional Design for Distance Education, Module 10, Evaluation. Online: http://www.usq.edu.au/dec/88031/M10.htm#Activity10.2

Biography

Nancee Simonson is a business journalist and training specialist, with a specialty in educational technology and web-based training design. She has designed and implemented an editorial training curriculum, internet training for editorial, and a computer journalism curriculum for the Bureau of National Affairs, Inc., an independent publisher in Washington, D.C. She has over twenty years experience in journalism and publishing management. Ms. Simonson is active in the arts and arts consulting, having served as a Commissioner on the D.C. Commission on the Arts and Humanities as well as various arts organization boards. She enjoys her two children and three dogs. Her educational accomplishments include a B.A. in Journalism (Marquette University), an M.A. in American Studies (George Washington University), a Fulbright research residency on international banking in Japan and an M.A. (Summer 1997) in Educational Technology Leadership (George Washington University). E-mail: nsimonson@geocities.com or nsimonson@bna.com.

The Bureau of National Affairs, Inc. is an employee-owned private business publisher in Washington, D.C., that publishes professional information in the business, legal and economic area. It is the largest independent publisher in Washington. BNA also produces business-related training materials.

Nancee Simonson

2108 Huidekoper Pl. N.W.

Washington, D.C. 20007

nsimonso@geocities.com

202-452-4329 (phone)

202-496-6902 (fax)

© Nancee Simonson, 1997

Authoring Tools for Courseware on WWW : the W^{3}Lessonware Project

Phil Siviter, School of Computing and Mathematical Sciences, University of Brighton, UK
Email Phil.Siviter@brighton.ac.uk

Abstract: This paper describes the W^{3}Lessonware project being carried out at the University of Brighton, England, the purpose of which is to create an integrated suite of tools for the production of world wide web (WWW) based educational materials. We introduce the project and its objectives, and give an overview of the main deliverables. We then outline the general strategy which was adopted to elicit an evolving set of requirements. We say a few words about how we integrated the main applications, and finally, we draw some conclusions regarding what the project has achieved and point out areas for future work, in particular the need for an evolving library of courseware oriented WWW templates.

Introduction

Overview

This paper summarises the W^{3}Lessonware project, an 18-month project managed by UKERNA on behalf of the Joint Information Systems Committee and carried out at the University of Brighton. The aims of the project were to produce a set of tools to facilitate the production of multimedia courseware based on HTML, and therefore deliverable directly on the world wide web (WWW). The official release of the tools can be downloaded from UKERNA's web site at http://www.tech.ukerna.ac.uk/ from which the reader can also obtain project documentation and the other deliverables from the project. The *latest versions* of the tools can be obtained from the W^{3}Lessonware web site at URL http://www.comp.it.brighton.ac.uk/w3lessonware/.

The project team comprised members with expertise in the development of WWW materials, educational and otherwise; the development of non-WWW courseware and courseware management tools; the development of computer based simulations of laboratory experiments, and network management and administration. Following an exercise in the construction of an example of WWW based lessonware in which we paid particular attention to monitoring our use of tools and techniques, and noting the problems which we encountered, we drew up the following broad set of requirements for a tool set aimed at facilitating the development of such materials. We felt that a WWW lessonware processor (W3LP) should include:

- a comprehensive, WYSIWYG, HTML editor which could incorporate a variety of media, in a variety of formats, into the HTML document
- an imagemap editor enabling the direct manipulation of hot regions and graphical elements on an image
- a utility to facilitate the structuring of large collections of documents
- a utility to facilitate the management of large collections of documents
- utilities to automate (at least partly) the tasks involved in making a large collection of HTML documents, and other files, look and behave like a coherent entity, e.g.,
  - managing the links between documents as documents are inserted into and removed from the collection
  - supporting the authors' view of the structure of the collection, and hence supporting meaningful operations on that structure (even though in reality the only structure is an arbitrary network of files)
  - supporting the presentation of these structures to the user, in the form of common navigational metaphors and their associated icons (next, previous, index, etc.)

We concluded that the above requirements fell into three main categories:

1. document editing
2. imagemap editing
3. structure editing

Accordingly, we proposed to build three core modules for W3LP - one to cope with each of these three areas. There would therefore be a document editor module (W^3HTMLEdit) an imagemap editor module (W^3MapEdit) and a structure editor module (W^3StructureEdit). Each of these modules would be able to communicate with the other two, as shown in [Fig. 1].

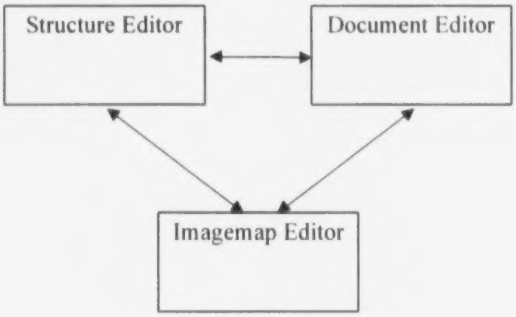


Figure 1: The three core modules of W3LP.

For example, the structure editor would present a view of the lessonware structure with icons representing files, and links between the icons representing links (e.g., HREFs) between the files. Double clicking on one of these icons would then open the relevant file in the document editor. The insertion or removal of links in the document, from within the document editor or the structure editor, would be reflected in the view from the other editor. Similarly, it would be possible to open the imagemap editor from within the structure editor. The structure editor enables the user to select a collection - any subset - of files from their project and pass their URLs to the imagemap editor, which will then create a user selected shape for each URL. The user can then position, size, and adorn the shapes to create a graphical menu whose hot regions hyperlink to the selected URLs. The imagemap editor can then converse with the document editor to enable the user to point to the position within an HTML file at which an imagemap should be placed. The document editor would be capable of displaying any in-line images (e.g., GIFs) within the document, in the same way that they would be seen when viewing the document with a typical web browser (e.g., Netscape, Internet Explorer, Mosaic).

In addition to developing the toolset, we created three examples of WWW based lessonware. These examples helped us to elicit requirements for the tools, and to test the tools. They also now act as templates for future lessonware developments. These examples can be viewed and downloaded from the aforementioned URLs.

Finally, we held two workshops during the course of the project - one in January 1996 and the other in July 1996. The second was multicast live over MBONE, and we hope to be retransmitting an edited version at regular intervals. The aims of these workshops were to get members of the courseware development community involved with the project, in order to:

publicise the project and its deliverables, especially the tools, so that they would become widely used

obtain feedback from the community regarding the tools' specifications, so that the tools would meet users' needs

The final specifications of the tools are lengthy documents and will not be repeated here. They can however be obtained from the UKERNA and Brighton web sites.

Choice of platform and development environment

The tools we have developed are all PC/Windows based, and were developed using the Borland Delphi environment (version 1.0). These platforms were chosen for the following reasons:

We decided that it was desirable to restrict ourselves to a single target platform so that we would make maximum progress with the tools, rather than spreading our effort across two or three platforms and not progressing as far.
The platform with the largest user base (by far) is PC/Windows.
Delphi was the preferred environment because it is based on an object-oriented language, and offered rapid application development facilities unrivalled, at that time, by any other object-oriented environment capable of producing stand alone executable files.

These decisions were the correct decisions in June 1995, when the project began. If we were starting the project today it is possible that we may make the same decisions again, although we would have to make a thorough evaluation of the latest Java development environments which have emerged recently to see if they truly compete with Delphi (and Borland's recently released C++ builder environment). If they do, then their cross platform promise would make them a very attractive proposition. In those circumstances the choice of hardware platform and operating system would be less crucial, although the PC/Windows platform remains the most cost effective, and popular.

General strategy

In this section, we describe the strategy we adopted for the execution of the project. Our approach was both iterative and collaborative. We planned the construction of three realistic examples of W^3Lessonware, for two main purposes:

1. to elicit and document tools requirements (or, to find out "the hard way" what tools would be useful)
2. to test the tools developed so far

The first example, produced at the start of the project, was primarily for purpose (1). The third example, produced near the end, was primarily for purpose (2). The 2nd example, produced midway through the project, served both purposes equally well.

From our experiences with the first example instance of W^3Lessonware, we defined a first tool set (project name: "toolset one"). We published a report on the W^3Lessonware example (available from the web site) and another on toolset one (also available from the web site). The idea was to get feedback very early on from the courseware development community. We wanted their ideas and opinions regarding tools requirements.

We then began specification and development of the core tools in the suite, regularly releasing updates to these tools on WWW, and asking for feedback from users.

Six months into the project, in January 1996, we held our first workshop (heavily oversubscribed) at which we presented the tools developed thus far, made useful personal contacts and received invaluable feedback and ideas. It is worth noting that the feedback received from this and the second workshop was greater in both quantity and quality than that which was received via WWW, email and mailbase over the entire duration of the project. Personal contact has proved to be more valuable than virtual contact - by several orders of magnitude.

Following the first workshop we entered our second major iteration, redefining the toolset (toolset two) and making substantial changes to the specifications of our core tools. A further six months work (which included the second and third W^3Lessonware examples) and WWW publishing brought us to our second workshop, in July 1996 (again heavily oversubscribed). This workshop was transmitted live over MBONE, and recorded in the University of Brighton TV studios. (We intend to re-transmit an edited version at regular intervals, both on MBONE and on the UK superJANET video network.) Once again we obtained invaluable positive feedback and constructive criticism from the delegates, which we acted upon in the next, and final, major iteration of the tools development.

The final period of the project saw us refining the tools to their current state of, if not perfection, then at least of professional standard and very real utility.

With hindsight we believe that this approach was basically correct, in that it enabled us:

to elicit external ideas and opinions, and to incorporate these into the tools

to be flexible in our specifications, as WWW developments occasionally threatened to overtake us

to make the user community aware of what we were doing, and thereby enhance the probability of the tools actually being used by a large number of people.

Integrating the Applications

The three main applications in the toolset (W^{3}HTMLEdit, W^{3}MapEdit and W^{3}StructureEdit) communicate with one another using the Microsoft Dynamic Data Exchange protocols (DDE). The HTML editor in particular acts as a DDE server to the other two applications, enabling the user to indicate simply, where within an HTML document a source anchor, or a destination anchor, or an imagemap, should be placed. (The imagemap editor also acts as a server to the structure editor during an operation which provides semi automation for the creation of graphical menus.) The details of the protocols used are documented in the HTML editor's on-line help. This information may be of use to other developers if they want to write applications to behave as clients in DDE conversations with W^{3}HTMLEdit, or if they want to write their own HTML editors to interact with the other W3LP tools in place of W^{3}HTMLEdit. (Such an editor must support the same protocols as W^{3}HTMLEdit in order to ensure correct operation of certain features of the structure and imagemap editors.)

In considering the requirements for inter-application communications, we did of course consider using OLE. However we eventually decided to use DDE instead for the following reasons:

We were developing for Windows 3.1 - a 16 bit platform - using Delphi 1.0, which produces 16 bit applications. Support for developing OLE clients is good in this environment, but support for developing OLE servers is poor. Placing OLE server capability in W^{3}HTMLEdit would have taken a long time and other aspects of the project would have suffered in the trade off.

The actual communication required to implement the above scenarios is very simple - DDE is perfectly adequate for this.

Should there be a further iteration of the tools, we would aim them at 32 bit environments, such as Windows 95 or NT, and develop them with Delphi 2.1, which has greater support for OLE server development. We would then use OLE to implement advanced features; for example, in-place activation of imagemaps from within the HTML editor, automatically invoking the imagemap editor (which would then be an OLE server).

Conclusions and Future Work

In terms of its original objectives, we consider the project to have been a success. The core toolset, comprising the structure editor, the imagemap editor and the HTML editor, presents users with an integrated suite of tools which makes creating lessonware for WWW a much easier process than it would otherwise be. The toolset is, as intended, of great use not only to the HTML novice, but also to the experienced user.

For example, a novice user can incorporate an imagemap into an HTML file without needing to know anything about <IMG> elements, USEMAP attributes, NAME attributes, shapes or co-ordinates. All they need to do is draw their hot regions on top of their image, specify URLs for each region, and point and click at the position within the HTML file at which the imagemap should appear.

Experienced users will also appreciate the time savings which the toolset can give them. For example, creating sequences of documents, and then editing those sequences by inserting and / or removing documents from the sequence, can be a very time consuming process requiring the editing of page counters in each document in the sequence, plus the editing of next/previous links (four links in three documents when inserting a new document into the sequence). The structure editor does all of this work for the developer, who simply drops the new document onto an existing sequential link.

The HTML editor enables developers to incorporate all of the major HTML elements, even if they do not know the syntax for those elements - they just supply the required attribute values (e.g., a destination URL, or the colour they want for their background) and the editor supplies all the mark-up. This not only eases the cognitive load on the developer, it also speeds up the whole process of writing HTML documents - even for experienced HTML writers. The built-in browser view also enables very fast switching between HTML and browser without the need to manually refresh / reload the document to see the effects of recent changes.

Naturally there are areas where we feel the tools could be improved. Any WWW project must face the fact that the WWW is evolving very rapidly indeed. Features which are commonplace now, such as Frames, Java, Javascript, ActiveX, did not exist when this project started. As these and other developments occurred we have had to decide whether and by how much, to modify the project's specific implementation goals in order to remain faithful to its original aims - of enabling the widespread exploitation of WWW for educational purposes. For example, we decided quite late in the project to incorporate some basic support for the creation and editing of frames within the structure editor. Time and resources are finite however, so some of the other planned features had to be set aside as a result. However, the frames feature has received very positive feedback wherever it has been shown, and overall we feel we have made the right compromise. We do of course intend to continue to develop and refine the toolset, as time and funding allows.

Thus we feel that the tools are a success, whether measured in terms of the project's own objectives, or in terms of how useful they are to anyone developing up to date, HTML-based resources.

The two workshops were also hugely successful, judging by the feedback we received from the delegates. In the second workshop especially, delegates who had never before used an imagemap (for example) were delighted at how easily they were able to generate one and incorporate it into their HTML (and the tools have been significantly improved since then to make it even easier). The workshops also provided an opportunity for personal contact between members of the courseware development community which we feel certain will prove fruitful in the future.

One area highlighted in the workshops - once again, the second workshop in particular - was the production and use of templates - ready made starting points for lessonware development which developers could use, and then tailor to their own purposes. The project has produced some templates for this purpose as planned, but although we always knew that templates were a significant resource, the feedback we have received has made us reassess the magnitude of this significance. We now believe that although sophisticated tools are an essential part of any developer's arsenal, the availability of a large, diverse collection of customisable W^3Lessonware templates, at a number of levels of abstraction (from a component on a single page to an entire course structure) would provide the greatest single boost in WWW lessonware development productivity.

To this end we are proposing a further project to develop such a library, along with tools for browsing the library, extracting and customising the templates, and putting them together to form coherent items of courseware for downloading to the developer's own site for completion. If our proposal is successful, we will look forward to building on the success of the W^3Lessonware project, to provide the community with a service consisting of a continuously evolving template resource base combined with courseware construction tools which are exceptionally easy to use (even by people with no technical knowledge of HTML). We hope that this will hasten the adoption of the WWW as a teaching and learning medium by an increasing number of education providers, and thus make a significant contribution to a more flexible educational provision for an increasing number of education consumers.

Cultural Impacts of the Internet and World Wide Web on a Computer-Literate Government Organization

Joe Sparmo, NASA/Goddard Space Flight Center, Email: joe.sparmo@gsfc.nasa.gov

David Matusow, NASA/Goddard Space Flight Center, Email: david.matusow@gsfc.nasa.gov

John Bristow, NASA/Goddard Space Flight Center, Email: john.bristow@gsfc.nasa.gov

Abstract: The Flight Dynamics Division (FDD) of the National Aeronautics and Space Administration (NASA) at Goddard Space Flight Center (GSFC) in Greenbelt, MD. USA recently conducted a study to gauge the impact of the rising influence of the Internet/Intranet on its staff. The FDD is a highly technical organization comprised of about 70 professional men and women who are considered highly capable of using computers at work. This paper merely describes this study and some of the results and observations that were derived. The authors do not propose that the results and conclusions drawn here have any significance beyond the organization that was studied.

To compile data for this study, the authors conducted interviews with a representative sample of people from the organization. The questions asked were the same for each interview and apart from the raw statistics of the answers, the authors also compiled additional insights from the interviewed pool. One of the most surprising findings is that the rise of the Internet/Intranet use in the FDD's computer literate business environment has had a notable impact on the FDD culture, despite the FDD's staff's familiarity with computer technology.

1.0 Purpose of the Study

For more than 25 years, the Flight Dynamics Division (FDD) at NASA/GSFC has been using computers for analysis of spacecraft data to perform mission planning, navigation support, and attitude determination. Until about 1989, the primary purpose of computers in the FDD was to develop and execute the large software programs used to perform these functions. However, similar to many organizations with high computer familiarity, in the late 1980's and early 1990's the professional community of the FDD first began its widespread use of electronic mail. Thus, desktop computing resources in the FDD also fed the demand for easy electronic mail access and other communication tools. By 1993, every desktop had a PC or a Macintosh and users were networked through various LAN configurations.

Unfortunately, the transition to desktop computing coincided with a tumultuous time within the FDD organization, and no social study was conducted. However, it was during this time that Internet/Intranet access in the FDD would evolve from a small amusement to a central mechanism; keeping the internal processes running and bringing the organization closer to the external world.

The accessibility of the Internet, the growing use of the World Wide Web, and the very evolution of the Information Age alone were not enough to fuel the expansion of the FDD's internal and external web domains. It took a combination of several committed champions and the willingness of management to embrace the new technologies. The champions would continue to demonstrate new benefits that Internet technology could offer the organization and persuade management to take some risks with these technologies. Once the FDD management started to take these risks, the only resistance was that of apathy and non-use which would not completely disappear until the use of various technology solutions were mandated.

The authors of this study are not attempting to use the results to push the FDD to change. Our purpose is merely to look at the cultural impacts of the Internet/Intranet and hopefully provide some useful commentary that other similar organizations may use to their benefit.

2.0 Methodology of the Study

The study would rely primarily on the collection of raw objective data along with some subjective measures that were captured by way of a series of 15 to 30 minute interviews conducted in October 1996. A total of 15 separate interviews were conducted in which the participants were asked a series of standardized questions. A panel of 3, which included the authors of this paper, were present at each interview. All 3 collected the data which was later synthesized together so that it could be studied for results and conclusions.

The interview was comprised of the following sections.

2.1 Introduction and background questions

It was important to clearly describe the purpose of the interview to each participant. The questions were not known to the participants prior to the interview and they were generally unclear as to the purpose of the study. The background questions were designed to capture some general characteristics from the participants such as the type of platform they preferred, if they used computers at home, and if they had any personal Internet access.

2.2 Personal Internet assessment

Each participant was asked to choose one of the following terms that best describes their relationship to the Internet (Apathetic, Enthusiastic, Competent, Resistor, or Challenged). This characterization would hopefully offer some insight when grouping the other responses together with those who had a similar self assessment.

2.3 Feedback on workplace processing mandates

As recently as August 1996, the FDD had developed some web-based tools that are accessed through an internal network (Intranet) using a browser like Netscape. These tools were designed to help automate standard administration processes such as requesting leave, filling out time sheets, and requesting work schedules. In order to insure that the full benefit of these tools were realized, the FDD management mandated their use by all employees. The study attempted to capture employee reaction to these mandates since it was relevant to impacts in the culture of the workplace.

2.4 General impacts

This section of the interview attempted to capture many different potential impacts by asking over a dozen short response questions. The hope was that some trends would emerge from the responses through correlation with other response data.

2.5 General questions

A few questions were asked to discover the preferences that the participants had for various Internet components over others. Also, the authors were looking for general feedback on the web-based administration tools described above as well as to offer the participants a final opportunity to comment subjectively on Internet impacts from their point of view.

3.0 Characteristics of Participants

The interview participants comprised a functional cross section of the people employed at the FDD. They are characterized as follows: manager, secretary, analyst, and developer. Each of the four categories of employee has a unique perspective of Intranet/Internet impact. All use computers every day in the FDD but, of course, for different purposes.

Managers in the FDD use computers to support their supervisory responsibilities. They make extensive use of word processing and spreadsheet tools to write memos and manage resources. They also heavily use their computers to communicate through electronic mail with others members of the Division as well as with many external people. The Intranet has allowed these managers to improve the day to day administration processes using the web-based tools.

Secretaries in the FDD also use the web-based administration processes. As time keepers, they receive all automatically generated electronic mail for processing bi-weekly time charges by the FDD staff. They communicate primarily through electronic mail, as that is the principal mechanism by which the larger Goddard organization disseminates information. The secretaries have also helped drive recent changes that have allowed the web to assist in replacing the many different forms that are typical to a bureaucratic organization. They do this by merely insisting that employees use the web-based forms.

Analysts and developers in the FDD comprise the majority of the professional people in the Division. They have dual roles of a singular purpose, to support the science objectives of the Agency in the area of spacecraft Flight Dynamics. Computers, of course, play a major part of this. For the analysts, the computer is a gigantic calculator that can be used to assist in processing large amounts of data to either graphically or tabularly provide needed information. The developers assist them by providing the software they need to accomplish these requirements. Their desktop computers not only provide support to these core objectives, but they also provide the electronic mail to communicate, the supporting software to develop code, execute programs, document creation, etc.

4.0 Findings

Correlation of answers from selected pairs of responses produced interesting findings.

4.1 Significantresults

The following is a collection of the study results that were of interest.

- All study participants preferred paper calendars over electronic.
- Those who claimed to be enthusiastic about the Internet also felt that the mandate to use the web-based administration tools was a Great Decision.
- Those who did not claim to be enthusiastic at best called the mandate an OK Decision. People who complained about physical problems associated with using computers generally were not enthusiastic.
- But they all concluded that the Internet makes their lives easier.
- For those who said that the Internet does not make their lives easier did say that they get too much electronic mail.
- Whereas, those who said that the Internet does make their lives easier also commented that they do not find that they spend too much time reading electronic mail and too little getting real work done.
- For those who said communications in general has improved thanks to the Internet also said that they have not reduced their physical interactions with others. They also say that their lives have improved and that they take the time to use proper netiquette.
- People who find themselves thinking of how to introduce new technologies do not complain about physical problems associated with computer use and they too strive to use proper netiquette.
- People who would like to see more Internet based automation do not complain about too much electronic mail.
- For the most part, people are either communicating at the same rate or more rather than less.

4.2 Comments on mandates

As mentioned above, the FDD mandated the use of the web-based administration tools in order to achieve their full benefits. The following is a summary of the verbatim comments received from this controversial item during the course of the interviews.

- "I'm surprised we needed a mandate, people should want to do it."
- "Mandates for administrative tasks are fine."

- "Mandates help overcome expected resistance, but in our environment (with all the computer knowledge) it should not be alot to ask."
- "Internet related mandates are like any mandate, they must have a solid reason for requiring their use."
- "Before diving headlong into a mandate, its important to show the benefits. If you find that 80% prefer using it then sure, a mandate to get the other 20% makes sense."
- "Mandates are usually bad, but for some certain circumstances, like when consistency is crucial, they should be levied. If a system is mandated for use, then if should at least be flexible."
- "Mandates could worry some people if they feel pressure to learn."
- "In a professional organization such as ours, it makes no sense that anyone would have a problem with any mandate."
- "The system must be mature before any mandate is levied."

4.3 Feedback on the web-based tools

These web-based tools that were mandated in the FDD have benefits and drawbacks that should be common to any organization that attempts to initiate a similar mechanism. Below is a breakdown of what the participants said.

Web-based tools benefits:

- easy access
- ease of communications (no need to track people down)
- makes sense, its more efficient
- gives immediate insight to what others are doing
- less paper in mailbox, more shelf space
- easy to use
- alot of up-to-date information with just a click
- saves time
- helps when searching for information
- helps address memory/storage problems

Web-based tools drawbacks:

- not complete, not consistent
- what if workstations are down
- always a work in progress
- inflexible
- creates resistance
- learning curve
- anominity
- less face to face interaction
- all must use to work
- secretaries/managers get alot more electronic mail
- fear that information will get lost

4.4 General comments

This remaining subsection presents the key observations provided subjectively from the participants.

- "The problem with doing things automated or on-line is that you lose the comfort of working in the physical word. With timecards or leave slips you could see it, and that was reassuring, but on-line, once you hit the button its gone and you hope the process works."
- "Social impacts of the Internet here at work can really be explained by a natural instinct to resist change. Its not a technical reason, its just human nature."

- "I still have not been able to figure out how to organize myself in an on-line world. I still end up printing off alot of electronic mail messages, especially for use in meetings, and to review (marking documents up with handwritten notes)."
- "The dream of the 80's was to have access to different computers from your desk. That wish has come true, but its obvious we did not think it all through. We are more productive but we are still as busy as ever."
- "I noticed many older people resistant to change. Most people resist change but eventually they get used to it and like it."
- "Can't find enough time in the day to learn all the things that are now possible, and that's frustrating."
- "There seems to be a need for more flexibility."
- "I would prefer not having to learn the Internet because I have enough to learn already, but I find that its impossible to ignore, and that annoys me."
- "As a secretary, I maintain the schedule of others, which means that the Internet has effected me, with on-line schedules I have to always be connected in order to get information that others request."
- "For personal use its great, as a phonebook, or some great reference tool."
- "Its only useful if most people use it, like with a newsgroup that my project uses, everyone has to use it in order for that to be an effective way to communicate."
- "As a college student I noticed that here at work people are more resistant to the Internet than at school. I guess its because people are older here and they are set in their ways."
- "I see an impact with incompatibility because of the diversity in desktop platforms, operating systems and software. But platform mandates would be a terrible mistake."
- "Less physical interaction is bad. (even though few claim to have reductions in their own physical interactions --- is this a myth?)"
- "Big impact in wasted time. Junk electronic mail, junk news in newsgroups, hard at times to find useful information."
- "Buy in is important. Time keepers are the best example, the web-based administration would have fallen flat if they did not embrace it."
- "As computer professionals, we see impacts because not all of society has changed, we still have to interface with the non-Internet world."
- "Low cost access opens the door for abuse. Users will get more electronic mail since its cheaper than sending (physical) junk mail."
- "Big impact with security. With the Internet, you don't need a degree in science or mathematics to cause havoc."
- "Very hard to review large documents, so I find I must get a large printout."

5.0 Conclusions

The raw data and the lists of subjective notes can be used and interpreted in several different ways. The authors have used various portions of these results to reach the following four conclusions. They felt that these were drawn from the most significant findings from the data.

5.1 The Impact of the Champions

People in the FDD are enthusiastic about the Internet and from those some champions emerge. As mentioned earlier, it is the champions who demonstrate the potential of the Internet in their organization. The study showed that there was a direct correlation between those who characterized themselves as enthusiastic about the Internet and those who felt that the web-based tool mandate was a great decision. Those who did not see themselves as enthusiastic about the Internet felt the mandate as an "OK decision", at best.

It can be concluded from this observation that those driving the changes of these new technologies realize quite clearly that in order to achieve the full potential of the Internet within an organization is for everyone to use it. That is why a mandate, which is generally viewed unfavorably in organizational surveys, was characterized as a great decision.

The impact that the Internet champions can have is only possible when management is willing to accept risk. But there is much at stake for the champions. If the technology that they champion does not demonstrate the benefits they propose then the larger organization may strongly reject it.

5.2 The Myth about Communication

A common criticism of the Internet is that it is creating a new type of isolation. This effect goes beyond the workplace to society itself. The feared impact is that a culture will emerge where people will not be required to even leave their homes in order to work and live a productive life. The fear is that this will create a breakdown in what is understood as communicating with others. Clearly the Internet allows more communication, but its the breakdown of face to face relationships that concerns many.

Assuming that the FDD is a typical workplace culture, this study shows this concern to be a mere myth. Not only do the participants insist that they are not communicating any less but they feel that their physical interactions with others in increasing as a result of the Internet. This is an interesting conclusion because today an employee in the FDD could do a good day's work without ever leaving their office. Despite this technological capability, they are interacting with more people than ever before (as a whole). The Internet is not creating a sense of isolation.

5.3 The Perception of Electronic Mail

There was a common theme in the response of those who did not have an enthusiastic perception of the Internet. These respondents did not feel that the Internet made their lives easier. The primary reason centered on their perception of electronic mail's impact on their working day. This group felt that they just spend too much time dealing with electronic mail messages. During a typical day, an FDD employee will receive between 20 and 60 messages.

It is interesting to point out that there is no correlation between the number of messages received daily and positive/negative Internet perceptions. Thus it can be concluded that Internet enthusiasm is tied to an ability to manage a typical day's worth of electronic mail. Some people know how to do it, and others do not. This is the underlying cause of raised frustration levels in the organization as it pertains to the Internet.

It is also important to discount some of the other perceived causes for low Internet enthusiasm. Some of the younger people in the organization felt that resistance was generated from older people who were just reluctant to change. Although its true that change is a factor that generates resistance, the fact that an employee was older did not correspond to Internet resistance; there was no trend that increasing age increased resistance. The younger people are generally more Internet inclined because of their recent college experience.

5.4 The Barriers to Organizing

Another interesting study finding for the FDD was that everyone preferred paper calendars over electronic calendars. The reason for this can be attributed to the lack of maturity in robust electronic calendars available to FDD employees, but it is also related to their comfort factor. This comfort factor is illustrated in some of the comments that referred to the Internet's lack of physical reassurances. A few years ago, employees would fill out a piece of paper to submit their time cards. This was a physical act, it could be filled out and seen, then handed off to another person to get paid. With Internet based forms replacing this mechanism, the employee relies on a virtual representation of the physical card and this, for some, is less reassuring.

Although Internet technologies are widely available, many in the FDD have yet to see true organizational benefits. Organizational in this context means organizing your workplace. Despite paperless administration processes, mandates to have all documentation on-line, and electronic mail and memos, many offices are still cluttered with documents and paper. Without a conscious effort, the rise of the Internet can make personal organization much more complex.

The final thought is this ... despite the high familiarity with computer technology in the FDD, it was discovered that impacts are indeed felt and are in many ways similar to those that would be found in organizations that have little computer experience.

Web-Based Requirements Analysis

Gees C. Stein - Faculty Associate, CAESAR, University of Texas at Arlington, USA, gstein@csr.uta.edu

James M. Mantock - Vice President, ScenPro, Inc., USA, jmantock@scenpro.com

Karan Harbison, Ph.D. - Director, Center for Advanced Engineering Systems And Research (CAESAR),
University of Texas at Arlington, USA, harbison@onramp.net

Stephen P. Hufnagel, Ph.D. - Director, CAESAR, University of Texas at Arlington, USA, hufnagel@csr.uta.edu

Abstract: A web-based search engine is described that uses relevancy measures to aid the requirements generation and analysis process. The search engine takes a requirement under consideration and generates an appropriate set of keywords for use in searching. These keywords are used to find relevant documents. Once the system finds relevant documents, the result is presented in an anchored HTML format so the user can view results of the search using a Web browser. By examining the relevant documents, the requirements development team can easily locate supporting or contra-indicating documents and other information.

Introduction

In this paper we describe a web-based search engine that uses relevancy measures to help requirements analysis teams with the following:

- Ensure consistency among proposed requirements
- Identify supporting information in related documents
- Identify inconsistencies with information in related documents

In the design and construction of complex systems, requirements development is a critical step. In a large system there can easily be thousands of requirements. During requirements analysis, one must ensure that the proposed requirements are consistent among themselves and do not duplicate one another.

In addition, there are often many other documents related to the requirements: standard policies and procedures, requirements documents from earlier systems, etc. While consistency between the proposed requirements and these documents is not mandatory, any inconsistencies need to be identified. In those cases where there is consistency, we have supporting evidence for the proposed requirements. In those cases where there are inconsistencies further analysis is needed. The inconsistency may be due to improvements the proposed system will make (a good inconsistency) or it might reflect a conflict.

A requirements *team* is usually formed so that the large volume of requirements and related documents can be developed and analyzed in a timely manner. It is also common for the team members to be geographically dispersed. All this strongly suggests a customized web search tool tied to a requirements repository.

For a discussion of requirements analysis as it relates to object oriented software development see [Booch and Grady 1997] and [Coad 1990].

Requirements and Relevancy

A subset of typical data fields for a requirement is shown in [Tab. 1]. In complex programs the requirements table contains many fields. Ideally, many of the requirements fields can be limited to a fixed set of alternatives.

This makes it much easier to maintain consistency of at least part of the requirements records. Because of all the specialized fields, the *Requirement Statement* field itself can be relatively simple.

Data Field Name	Description	Example
Requirement Statement	Free Text	System shall provide for storage of medical training records
Functional Area	Gross Classification of Requirement, i.e., Threat Assessment, Training…	Training
Information Product	The Information Product Associated with the Requirement	Training Attendance Record
Data Elements	The Data Elements in the Information Product	Attendee, Class Name, Date, Instructor, Location, Pass/Fail
Flow	Where the Information Product Comes from and Where it Goes	Instructor sends info to Medical Officer after class. Medical Officer stores info.
Activity	The Activity Associated with the Info Product, i.e., Generation, Storage…	Storage
Justification	Reasoning Behind Proposing this Requirement	Ensure current records of medical qualifications are maintained.
Phasing	When the Requirement Is Active, i.e., Deployment, All Phases…	All phases
Current Method	Description of How the Requirement is Currently Met	Transmission of info is done via e-mail. Info is transcribed onto Form XYZ and filed.
Job Classification	The Job Class. of the Person Currently Fulfilling the Requirement	Medical Officer
Location	Where the Requirement is Fulfilled, i.e., Headquarters, Field Station…	Regional Support Office
Timing	Any Periodicity Associated with Fulfilling the Requirement	Within 5 days after completion of class.

Table 1. Requirements Fields and Example Requirement

Depending on the nature of the requirement, the contents of any or all of the requirements data fields may be needed to identify consistencies and inconsistencies.

We propose a search engine to find information *relevant* to a given requirement in documents. Relevancy isn't concerned about consistency or inconsistency. The requirements team wants to see both. Relevant information that is consistent supports the requirement. Relevant information that is inconsistent needs further examination.

Considering the large volumes of potentially relevant material, automation is almost mandatory. Modern search engine technology provides the foundation to do the comparisons quickly, but the current search engines do not provide the functionality needed for requirements analysis.

Another advantage of using search engine technology is the ability to set anchors into particularly relevant paragraphs. Documents can be of considerable length: when retrieved it is often difficult to locate the paragraphs of interest. Paragraph anchoring is expected to substantially improve productivity of analysts.

Proposed Web Tool

To find documents relevant to a given requirement, we adopted a keyword search approach. First, we use the requirement to generate an appropriate set of keywords for use in searching. Second, these keywords are used to find the relevant documents. Third, once the system finds relevant documents, the result is presented in an HTML format so the user can view results of the search using his/her Web browser. A variation on this using hypertext can be found in [Perlman 1989].

Keyword Generation

The documents we are interested in could restate the requirement using different wording, as a generalization of the requirement, as a more specific description of the requirement, or it could contradict the requirement on one or more details. All these documents are considered relevant.

In order to find these documents, one cannot just use keywords that are part of the requirement. First, the most meaningful words of the requirement are extracted. Currently, we determine 'meaningful keywords' by hand, based on a frequency count over all requirements. Later this method has to be refined. This initial set of keywords is expanded in several ways—by adding synonyms, antonyms, hyponyms, specific instances and inflections. (Note we could stem the words in the searched text to their root form, instead of giving all possible inflexions of each keyword. This will be explored in the future.)

This expanded set of words is then used by the search engine. Each word has a 'weight' attached to it: a number indicating its importance. The search engine makes use of this when judging the relevancy of a document. How the weights get assigned is described in 'Future Work'.

The search engine uses a set of keywords (the expanded keyword set) along with their weights to find the relevant documents. The search engine is described in detail in the next section.

The Search Mechanism

The search mechanism takes a set of keywords together with their weights as input. Typically the set of documents searched over is a constrained set selected by the requirements team. For each document the following happens. The system finds those words in the document which matches a keyword. Each match represents a number: the weight associated with the keyword. As a default this weight is 1, but for more important keywords the weight can be higher, typically 2. The search mechanism then calculates at a *paragraph* level the total of the weights for the matches found. This is the paragraph's score and is an indication of the *relevancy* of this paragraph to the requirement. The user can specify a threshold, i.e., a minimum score, that a paragraph must meet to be considered relevant. Only those paragraphs are shown to the user.

Our search engine has the following characteristics:

1. Searches use a relatively large set of keywords
2. Keywords have weights (numbers) attached to them
3. Keywords are connected only via weighted 'OR' connectives
4. Relevant documents must contain a certain number of keywords (set by the user)
5. Keywords should be 'close' to each other; a large document that contains several of the keywords, but all at considerable distance from each other, is unlikely to be of interest. Keywords should preferably be found in the context of some of the other keywords. Often words have several senses; looking for a word in a certain context (namely a context containing several of the other keywords) makes it more likely that the right sense of the word is found. 'Close' is a subjective notion: in our approach we consider words to be close if they are in the same paragraph
6. What part of the document is relevant is identified. In our case it does this at the paragraph level
7. Relevant text is scored
8. Geographically separated users are supported since it is Web-based

Existing search engines have been considered, but were found to be insufficient in one or more of areas mentioned above (many comparative studies of search engines can be found, for instance [Venditto 1996]).

A brief review of some of the most popular search engines applied to requirements analysis follows:

Alta Vista http://altavista.digital.com/
Alta Vista provides a full support for boolean operators in queries, but does not perform concept-based searches. This makes it unnecessary to type the expended word set. Retrieved documents are ranked, but hits are not anchored and there is no indication of how closely the match is.

Glimpse http://glimpse.cs.arizona.edu/
Glimpse anchors text, and one can configure it search on a paragraph level. Glimpse has to be given the expanded, disjunctive, keyword set. However, Glimpse cannot handle many keywords: it becomes extremely slow (or does not respond at all). It does not rank the output.

Excite http://www.excite.com/
Excite claims to be the only search engine that is smart enough to have some notion of 'related' concepts. In their example, they say that the search engine will realize that 'dog care' and 'pet grooming' are related topics. This makes it unnecessary to type the expended word set. However, this makes the search engine also rather mysterious, in the sense that it is not always clear why a document (that has a high score) was retrieved.

The search is done for a disjunctive set of words. Excite does rank the retrieved documents (it gives a percentage). It does not anchor text. Since it is fed an 'or' list of keywords, it returns a relatively large portion of the documents. We do not know if the score takes 'closeness' of the key 'concepts' into account.

Infoseek http://www.infoseek.com/
It is not clear whether Infoseek automatically searches for synonyms, etc. so an expanded word-set has to be given. The search is done for a disjunctive word-set. Infoseek does rank results, using percentages. It does not take 'closeness' of keywords into account. The text is not anchored. It is possible to search in a particular site.

Lycos http://www.lycos.com/
One can give Lycos fragments of words, to do a partial match. For instance, if one gives 'gard$', the system will match this with words that start with 'gard': garden, gardens, etc. This helps, but is not enough when one also needs synonyms, hyponyms etc.

Lycos does ranking: it gives a percentage and a number saying something about the keywords found, for instance, "Ranking: 100%, 5 of 13 terms". Lycos does not anchor the text. It is not clear whether the ranking takes the 'closeness' of the keywords into account. One can vary the matching from "loose match" to "strong match". Given the same keywords, the first option returns many more documents than the last. It is not clear how the search changes when changing this search option.

WebCrawler http://www.webcrawler.com/
WebCrawler does not anchor the text, and synonyms, antonyms etc. have to be input by hand. It does rank the documents found. It is possible to explicitly search for two words that have to be within a certain word distance of each other, but the general 'closeness' as defined earlier is not possible, nor does WebCrawler's ranking take closeness into account.

Yahoo http://www.yahoo.com/
Yahoo does not anchor the text, nor does it rank the documents found. It does give a classification of the documents, at a high level. It was not clear whether Yahoo would look for inflections, so an expanded word is required.

For none of the tools described above is it possible to assign a threshold: a minimum score a document should have in order to be returned to the viewer.

A general problem with the commercially available web searchers is that it is very difficult to find out exactly how the search is carried out. In addition, it is often impossible to restrict searches to directories specified by the user. Indices are often generated once a week, and might thus miss newly added documents.

Current State of Work and Future Work

Currently, we have a working system that meets the requirements described above. However, as the system changes and the amount of documents becomes larger (which is to be expected), speed will become a crucial point, and special measures will be taken to maintain reasonable response time.

Our next step will be working on deriving a good keyword set automatically from a given requirement. We have to decide which fields should be used for this, and, given a field, which words will give the best results. Also, the chosen keyword set has to be expanded automatically. We plan on a user interactive system where the user can adjust weights of the keywords.

In the future we hope to add some learning capabilities to the system by using previous evaluations of retrieved text units (paragraphs) by the user. Similar work has been done, among others, by Pazzani [Pazzani, Muramatsu and Billsus 1995], but on a document level for less refined topics like 'BioSciences' or 'Music' [Armstrong, Freitag, Joachinms and Mitchell 1995], also at a document level.

We also intend to work on a tool to extract the (partial) requirement described in the retrieved text, in order to compare it with the initial requirement used to search for relevant documents. In this way we hope to develop a tool that supports the user in finding relevant documents and in deciding how a document relates to a requirement.

Conclusion

We believe the problem of finding relevant documents for requirements analysis is a general problem that is of interest to many organizations. To support this task we are developing a tool that automatically generates a set of keywords to support locating relevant documents. We have developed a prototype search engine that uses these keywords to not only find relevant documents, but score and anchor relevant paragraphs. A web-based implementation was chosen to ensure geographically dispersed teams would be supported.

Acknowledgements

This work is supported by NRaD contract N66001-94-D-6054.

References

[Armstrong, Freitag, Joachinms and Mitchell 1995] Armstrong, Robert, Dayne Freitag, Thorsten Joachims and Tom Mitchell (1995). WebWatcher: A Learning Apprentice for the World Wide Web. In *Working Notes of the AAAI Spring Symposium Series on Information Gathering from Distributed, Heterogeneous Environments*. Menlo Park, CA: AAAI.

[Booch and Grady 1997] Booch, and Grady (1997). *Object Oriented Analysis & Design with Applications*. Addison-Wesley.

[Coad 1990] Coad, Peter (1990). *Object Oriented Analysis*. Prentice – Hall.

[Pazzani, Muramatsu and Billsus 1995] Pazzani, Michael , Jack Muramatsu & Daniel Billsus (1995). *Syskill & Webert: Identifying interesting web sites*. http://www.ics.uci.edu/~pazzani/Syskill.html

[Perlman 1989] Perlman, G. (1989). System design and evaluation with hypertext checklists. In Proceedings of *1989 IEEE International Conference on Systems, Man and Cybernetics*. p. 1187-93 , Vol 3. NY, NY: IEEE

[Venditto 1996] Venditto, Gus (1996). Search Engine Showdown. IW Labs tests seven Internet search tools. *Internet World*, (7:5) p. 79-86. http://www.internetworld.com/1996/05/showdown.html

Copyright 1997. Association for the Advancement of Computing in Education (AACE).
Distributed via the Web by permission of AACE.

A 3D Topographic Map Viewer for the USA

S. Augustine Su
Department of Computer Science
University of Maryland, College Park, Maryland, USA
Email: su@cs.umd.edu

Shueh-Cheng Hu and Richard Furuta
Center for the Study of Digital Libraries
& Department of Computer Science
Texas A&M University, College Station, Texas 77843-3112 USA
Email: shuehu@cs.tamu.edu, furuta@cs.tamu.edu

Abstract: Traditional maps are abstractions of the real world. Among them, topographic maps are the most self-explaining way to make clear the correspondence of physical and man-made characteristics. However, topographic maps consisting of contour lines are not intuitive. In this paper, we present a Web 3D Topographic Map Viewer that generates 3D maps in VRML 2.0 on demand for areas in the contiguous USA, based on the U.S. Geological Survey's Digital Elevation Models. This provides immediate access to a huge elevation database and facile visualization of terrain.

Introduction

Maps that convey geographical information are ubiquitous on the World-Wide Web. In 1993, a Web server, the Xerox PARC Map Viewer, began to dynamically create GIF images of maps for user-selected areas in the world using a geographic database based on Digital Line Graph (DLG) data [Putz 1994]. Borrowing ideas from that Web site, another Web server, the U.S. Census Bureau's Tiger Mapping Service, set out to provide U.S. street-level maps in late 1994, based on TIGER/Line data [US Census Web]. Since then, Web servers that create maps on demand to serve various needs have been very popular.

In 1995, the Virtual Reality Modeling Language (VRML) 1.0 was announced; a year later, VRML 2.0 was finalized [SDSC Web][SGI Web]. Learning the intrinsic power of navigating three-dimensional (3D) worlds from any VRML browser or plug-in, we had a thought that all two-dimensional traditional maps could be transformed into 3D counterparts in VRML so that users could get intuitive feelings of 3D terrain by viewing them from various viewpoints. This would facilitate learning the relationship between topography and the information carried on 2D maps.

For that reason, we started to build 3D terrain in VRML 2.0 using public-domain geographical data, specifically the U.S. Geological Survey's (USGS) 1-degree Digital Elevation Models (DEM) [USGS Web]. At first, topographic maps of the Washington, D.C. area and of Texas were built [Su, Hu, & Furuta 1997], where we presented an example about how the maps could be used in the teaching of a high school geography class.

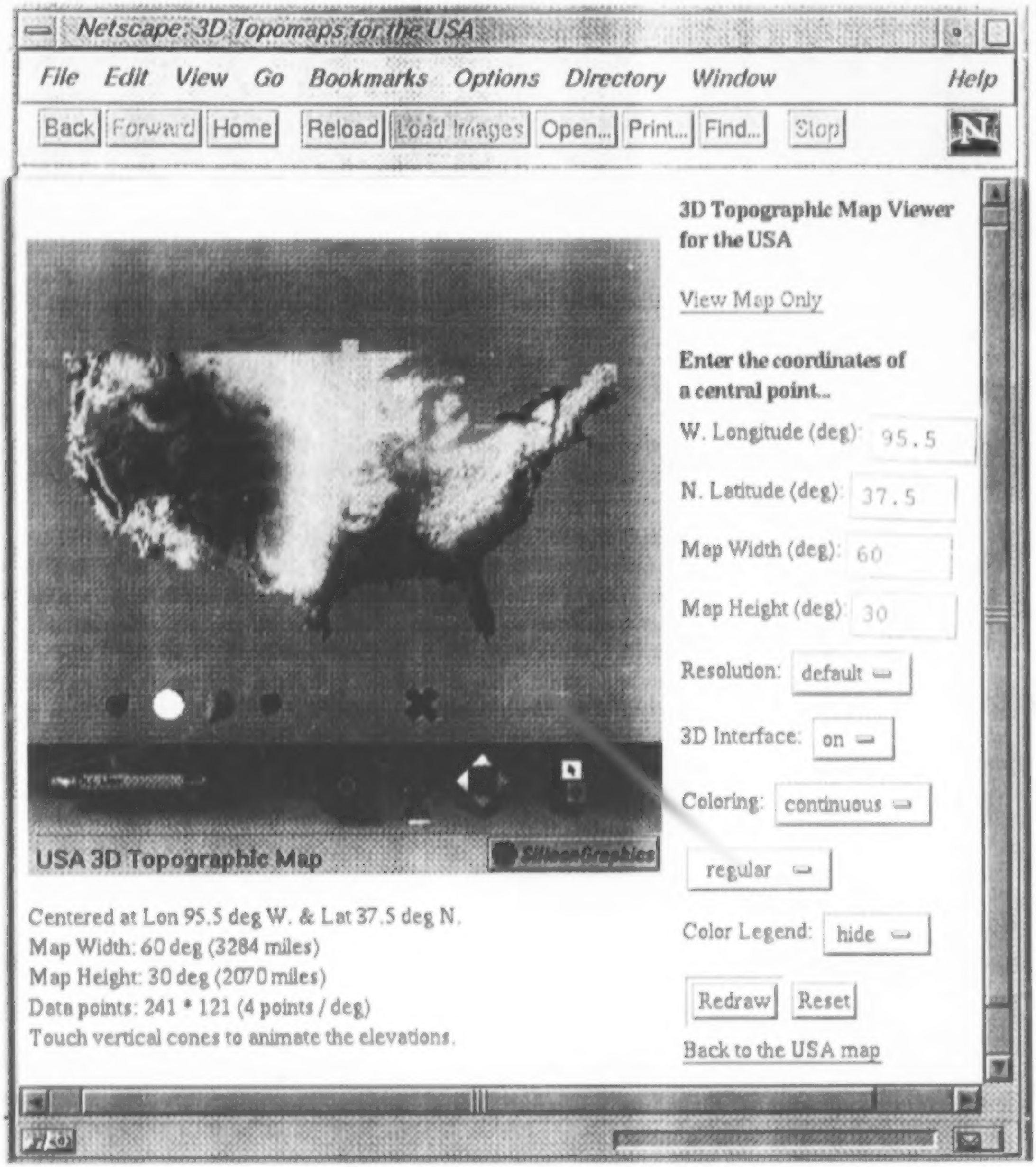
In this paper, we will report the work of building 3D topographic maps in VRML 2.0 for the U.S., excluding Alaska and Hawaii. Since VRML documents having rich contents are usually large, they take a very long time to download from the Web, and later overdrive or even cause today's VRML browsers or plug-ins to hang. Therefore, we built a Web site, the 3D Topographic Map Viewer, to assist viewing 3D topographic maps by limiting the size of data transferred on a user's request. The DEM data that we use has the resolution of 1,200 elevation data points per degree (within 300 feet has a datum point), so our 3D Topomap Viewer achieves this level of resolution.

WebNet 97, Toronto, Canada (Oct 31-Nov 5, 1997).

The first author currently is visiting the Center for the Study of Digital Libraries at the Texas A&M University.

Traditional topographic maps use contour lines to represent elevations. This is not very intuitive, but the elevation can be represented accurately. Other topographic maps use a small number of colors to represent elevations. This strengthens the human's perception of areas with the same elevations, however, the distinctness of different elevations depends on the number of colors used. We have experimented with various color schemes with a continuous color spectrum in assisting terrain visualization.

In the following section we will demonstrate two examples using the Topomap Viewer. We will then describe its user interfaces and provide more detailed information about building the Web site.


Netscape: 3D Topomaps for the USA
File
Edit
View
Go
Bookmarks
Options
Directory
Window
Help
Back
Forward
Home
Reload
Load Images
Open...
Print...
Find...
Stop
USA 3D Topographic Map
SiliconGraphics
View Map Only
Redraw
Reset
Back to the USA map

3D Topographic Map Viewer for the USA

Enter the coordinates of a central point...

W. Longitude (deg): 95.5

N. Latitude (deg): 37.5

Map Width (deg): 60

Map Height (deg): 30

Resolution: default

3D Interface: on

Coloring: continuous

regular

Color Legend: hide

Centered at Lon 95.5 deg W. & Lat 37.5 deg N.
Map Width: 60 deg (3284 miles)
Map Height: 30 deg (2070 miles)
Data points: 241 * 121 (4 points / deg)
Touch vertical cones to animate the elevations.

Figure 1: The USA Topomap Viewer

BEST COPY AVAILABLE

Viewing USA Topography on the Web

[Fig 1] shows the initial display encountered by users of our 3D Topomap Viewer. On the right is an HTML (HyperText Markup Language) form that allows users to specify an area of request and its settings. The VRML plug-in on the left shows a map of the contiguous USA. There are two sets of controls within every 3D map: one set is for navigation; the other for elevation exaggeration, both of which will be discussed later.

Supposing a tourist wants to see the terrain of the Yellowstone National Park. After consulting a guide, he/she can fill in the form on the Topomap Viewer with numbers: west longitude at 110.5 degrees, north latitude at 44.5 degrees, width 1.8 degrees, and height 1.6 degrees. A map of the park will be generated when the "Redraw" button is clicked. The map mostly is colored in wood and dark wood because the elevations in the park occur in a relatively narrow range when compared to the USA as a whole (see the color legend discussion below). If the user wishes to increase the number of colors shown on the map, he/she can select the "optimized" color mode rather than the "regular" mode in the form and click on "Redraw" again. Subsequently, a new map with more colors will appear. Standard topographic maps exaggerate the display of elevation to highlight differences. Our user can do this also, selecting the degree of exaggeration by touching the control cones in the VRML world to animate the changing of elevations. Finally, by dragging the trackball, thumbwheel, and pan control in the plug-in, a desired viewpoint will be reached such as in [Fig. 2], where Yellowstone Lake is in the center and is colored in orange to show its elevation. If the user desires to see a bigger map than the default size of 400x400 pixels, he/she can click on the Hypertext link "View Map Only", switching the display to one in which the map will occupy the whole display area of the browser. If the user wants to go back to the starting point, he/she can click on the "Back to the USA map" link to get back to the top-level USA map. Choosing the "higher" resolution in the form results in decreasing the fuzziness of the map display with a higher data density on the map at the expense of a significant increase in data transfer and consequently in graphics rendering time.

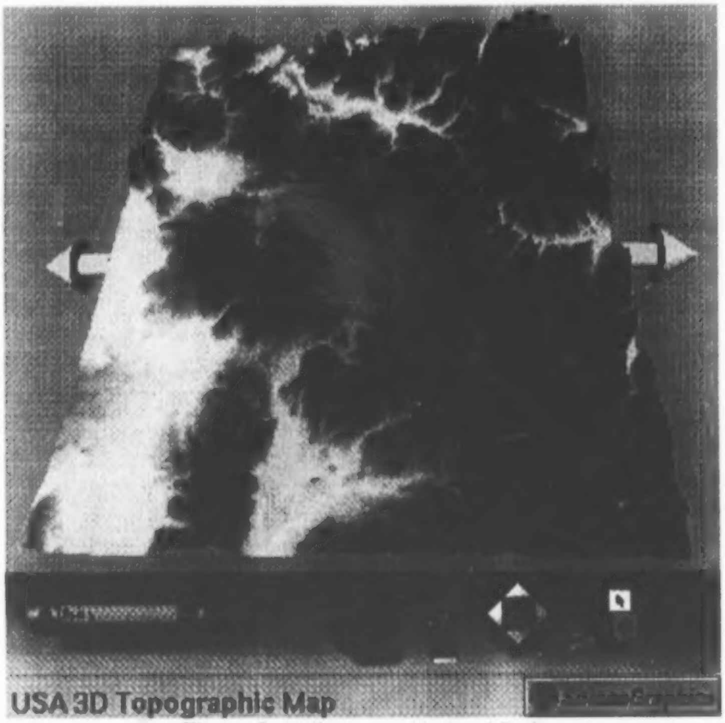


Figure 2: Yellowstone National Park

BEST COPY AVAILABLE

As a second example, we consider a user who wants to find Seattle. Here, navigation will be purely through the controls on the map, without the use of the HTML form. For people familiar with the US geography, it is not too hard to find out that the city is located in the corner of the contiguous US. So, the user can click on any point in the northwest part of our US map, and then click on the blue X-shaped control in the VRML world to initiate a zoom-in. Or, alternatively, the user can click on one of the yellow arrows surrounding the map in the VRML world to pan until the target is reached. The zoom-in and pan can be repeated to narrow down the searching area to locate the city. As before, the viewpoint and elevation exaggeration can be adjusted. The result is shown in [Fig. 3], where Seattle appears in the center of the map, flanked by Puget Sound and the Olympic National Park to the west and by the Cascades Mountains to the east.

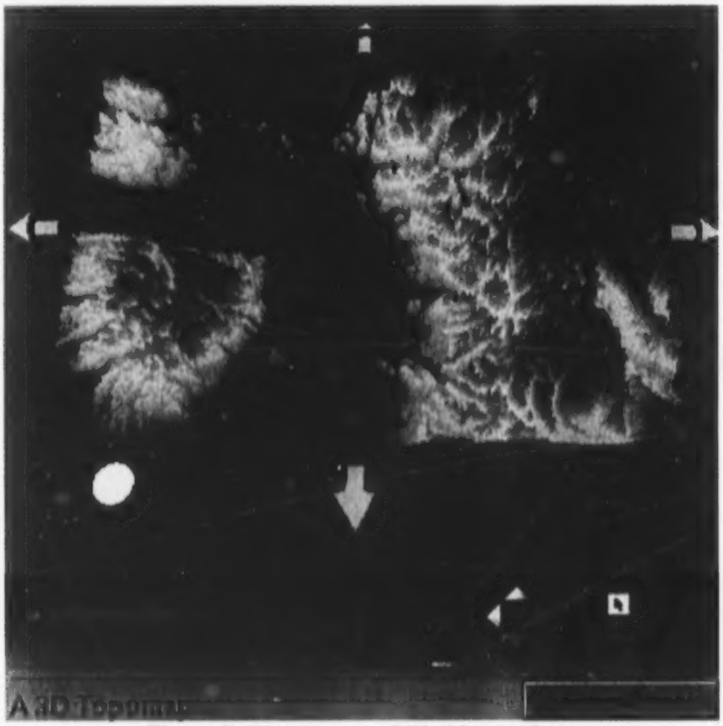


Figure 3: Western part of Washington State

User Interfaces

The user interfaces of the Topomap Viewer were designed to be simple and intuitive. However, each component of the interface still deserves explanation.

When 3D maps first are brought up, the elevation is not exaggerated, i.e., the elevation and the ground are at the same scale. As noted before, exaggeration is used commonly in topographic maps to highlight relatively slight ups and downs of terrain found in large-area maps. To the lower left of our maps, there are four vertical cones that are used together to adjust the scale of elevation exaggeration: one small and one large pointing-up cones to scale up the elevation, and one small and one large pointing-down cones to scale down the elevation. The large cones will animate the elevation changing in a larger scale than the small cones. The changing rates depend on the size of maps.

To the lower right of maps, there are three blue controls corresponding to three actions: the X-shaped one is for zoom-in, the diamond-shaped for pan, and the cross-shaped for zoom-out. The user has to click on the map to set a central point, and then click on one of the blue controls to start the corresponding action. If no central

BEST COPY AVAILABLE

point is selected, the old central point is assumed. For the top-level maps, only the zoom-in control exists since the others are not necessary. The zoom in and zoom out factors are fixed at 2.

Except for the top-level maps, four yellow control arrows for the four directions (east, west, north and south) exist adjacent to the four sides of maps. The map is panned when a directional arrow is clicked. The new view keeps the previous width, height, and display settings.

There are eight scales of map resolution provided: 4, 10, 20, 40, 100, 200, 400, and 1200 data points per degree. However, the user is only allowed to choose among three options: the "default", "higher", or "lower" resolution ("default" is automatically picked by the system from the eight available scales). The elevation data are uniformly sampled from the geographical database.

As for the color legend, each elevation has its own color on a map. Nine major colors were selected for the elevations above the sea level: dark green, green, light green, yellow, orange, red, wood, dark wood, and white (from low elevations to high). In the "continuous" mode, the color change is continuous: one major color, and then colors interpolated in RGB space between that major color and the following major color, and then repeated from the above list from low to high. The "discrete" mode, however, only uses the nine major colors to specify nine elevation ranges. In the "regular" mode, a global elevation color legend is used for all areas; however, in the "optimized" mode, a localized elevation color legend will be used to adapt to the elevations in the requested area.

Implementation

The HTML document composing the Topomap Viewer is dynamically generated by a CGI (Common Gateway Interface) script on the server, whenever the user makes a connection to the Web site. In this HTML document, there is an important URL (Uniform Resource Locator) within the EMBED tag for the VRML plug-in, which is used to initiate another CGI script on the server to generate a VRML document dynamically. To compose elevation information in the VRML document, this CGI script calls a separate program to retrieve data from the geographic database.

Parameters are passed to the server via the Fill-In Form and query strings that are attached to the scripts' URLs. The top-level VRML documents are cached instead of dynamically generated. The two CGI scripts were written in Perl and the program was written in C.

Our database is based on USGS 1-degree DEM data. Each DEM is an ASCII file containing 1,200x1,200 data points in integral meters which record the elevation uniformly among a 1-degree-by-1-degree block. The size of a DEM is 9.8 megabytes. After GZIP compression, the size is generally in the range of 0.5 to 1.5 megabytes. There are 935 blocks in total for the US lower 48 states. Each DEM is named by place and direction, such as "San Antonio West". We created a binary file for each DEM, but named it by its own longitude and latitude. To shrink the size of the database, our files use various numbers of bits to encode elevation data depending on the maximum elevation in the files. In the result, we have a database of 1,200x1,200x935 data points, and the size is 1.55 gigabytes, which is about the size of the original GZIP-compressed DEM data. The advantage of our approach is that given a longitude and a latitude, the random access of its elevation datum is made efficient.

To expedite generating elevation information for VRML documents, the original database was uniformly sampled to create two lower-resolution databases: one is 200 data points per degree (the size is 43.6 megabytes); the other 20 data points per degree (974 kilobytes). In other words, one of the three databases will be used depending on the resolution of maps.

We tried our best to limit the size of each VRML document into a few hundred kilobytes (before compression) when the "default" resolution is selected in the form. The resolution used is a function of the map size.

Our scripts were written to be suited for currently the most popular VRML 2.0 browser or plug-in, SGI Cosmo Player [SGI Web]. The coordination among the four elevation control cones is done by a small script written in VRMLScript, a subset of JavaScript, as is the control of clickable zoom-in, zoom-out, and pan.

Discussion and Future Work

Composing 3D maps is a long process of trial-and-error. At an early stage, we tried several versions of 3D maps using the LOD (Level-of-Detail) node of VRML 2.0. But none of them can work without problems due to the reason that the current implementation of the LOD node does not swap data out when the data are not needed in the current viewpoint. Consequently as more and more data are downloaded by LOD nodes, the chances increase that the VRML browser will run out of swap space and crash. Therefore, we chose to use a map viewer rather than a single huge LOD hierarchy for the USA map. This allows map viewing not only on high-end SGI workstations, but also on low-end PCs.

Since the DEM data do not provide any boundary information, some low elevations in the maps are inevitably "washed out" because elevations below 0.5 meter are rounded to zero in the DEM data (therefore, treated as the ocean in our maps). This unfortunately compromises the accuracy of coastal lines.

Our Topomap Viewer is still not a full-fledged map viewer: the maps are lacking political boundaries such as state and county lines, city information, and water bodies. Future improvements may also include a user-specified elevation color legend: users are allowed to manipulate multiple thumbs upon a scrollbar to specify corresponding elevations for major colors. The use of color schemes to visualize terrain is only one of many possible ways to do so. Finally, selection by city names and zip codes may be useful to avoid the need to consult an external reference when locating an area.

Conclusion

Traditional 2D maps are abstractions of the real world. Maps built in the 3D models have the potential to turn the abstractions back into objects that we can feel intuitively and strongly. The Topomap Viewer is our first investigation of this kind, and provides easy access to a huge elevation geographical database at a US national scale.

Our Web site for the 3D Topomap Viewer is at http://www.csdl.tamu.edu/topomaps.

References

[Hartman & Wernecke 1996] Hartman, J. & Wernecke, J. (1996). The VRML 2.0 Handbook: Building Moving Worlds on the Web. Addison-Wesley.

[Putz 1994] Putz, S. (1994). Interactive Information Services Using World-Wide Web Hypertext. *World-Wide Web (W3)*. Geneva, Switzerland, May 1994. Elsevier Science. Also available at http://pigeon.elsevier.nl/cgi-bin/ID/WWW94. The Xerox Map Viewer is at http://pubweb.parc.xerox.com/map.

[SDSC Web] The VRML Repository, a Web site at http://www.sdsc.edu/vrml/.

[SGI Web] vrml.sgi.com, a Web site at http://vrml.sgi.com.

[Su, Hu, & Furuta 1997] Su, S.A., Hu, S.-C., & Furuta, R. (1997). VRML-Based 3D Topographic Maps: Enhancing Educational Applications. *Educational Multimedia & Hypermedia (ED-MEDIA) and Educational Telecommunications*

(ED-TELECOM). University of Calgary, Calgary, Canada, 1997. Association for the Advancement of Computing in Education.

[US Census Web] Census TMS Home Page, a Web site at http://tiger.census.gov.

[USGS Web] USGS Geo Data, a Web site at http://edcwww.cr.usgs.gov/doc/edchome/ndcdb/ndcdb.html.

Acknowledgments

This material is based in part on work supported by the Texas Advanced Research Program under Grant Number 99903-230.

Automatic Interests Extraction Chasing the Browsing and Event History

Seiji SUSAKI, Tatsuya MURAMOTO
NTT Information and Communication Systems Labs.
1-1 Hikarinooka Yokosuka-shi Kanagawa, 239, Japan
E-mail : {suzaki, muramoto}@isl.ntt.co.jp

Abstract : Many collaborative information systems called social filtering systems have been developed recently, which WWW (World Wide Web) users are directed to useful information by other users, because there are a lot of resources on the WWW and it is hard to specify which information is important for each user. Although these systems are efficient for users about advising useful comments that others store in their database, there are a lot of problems making use of this information. This paper first summarizes the social filtering system and problems, and describes a framework for extracting user interests automatically for application in a social filtering system whose features are : (i) not to impose a cognitive overhead, and (ii) to acquire user interests automatically. Using this system, it can specify to the noteworthy user information in which one is interested, and we clarify the possibility of realizing to make social filtering system.

1. Introduction

Users of the WWW are bewildered by its rapid growth, so they are seeking help with browsing, usually in the form of directory service or a robot-based search system. TITAN (Total Information Traverse AgeNt [Susaki 96]) is a robot-based search system that assists users by meaning of its cross-lingual search support features. Although such systems are very useful, they can not fulfill their original functions now. For example, say one wants information on Artificial Intelligence (AI) and attempts to retrieve that information from very large database using a robot-based search. There will be many results from their query, so one has to sift through them to find ones which answer his query. If there are too many results on the screen, choosing which pages to visit next can be confusing. This suggests the need for new tools to assist in making these choices. For this purpose, social filtering systems [Malone 87] have been developed recently. Our goal is to create a useful system using this framework for users to help each other. To achieve this goal, we first try to identify user interests automatically to determine the values of pages. The next section describes the framework of the social filtering system, and the features used to construct these systems. After that, we explain our prototype, which considers the users action history to follow his activities. Finally, we evaluate our method ability to identify user interests, and conclude.

2. Social Filtering System

The main idea of a social filtering system is for each user to share information they obtained. In the real world, we are sometimes told where a important information is and to which information we should refer. It is especially efficient for users interested in the same topics to suggest information to each other. Suppose someone interested in AI gathers some information using a WWW browser such as Netscape Navigator or Internet Explorer. He or she evaluates how useful this information is, learning which information is meaningful in relation to AI. In the social filtering system, this knowledge is passed on others with same interests. The procedures used in social filtering systems are as follows.

- Evaluation
  Browsed pages must be evaluated. The evaluation can be conscious or unconscious. In conscious evaluation, when one visits a page, an evaluation window pops up where they have to input their opinion [Firefly 96]. The evaluation of the target page is stored and used for social filtering. The unconscious approach makes use of the number of times the target page is visited [Res[illegible]ck 94]. It is considered that this number corresponds to the level of user interest.
- Classifying
  Users can be classified according to their interests. The method makes use of the user evaluations of pages. Users belonging to the same community are interested in similar pages, so they can be

classified according to tendency in their referring history. For suggesting information to users, it is necessary to identify whether a user has the same interests or not. This is determined by calculating the correlation between other user references and this user's profile [Shardanand 95].

- Suggesting
  After identifying an appropriate interest group, this system suggests useful information to users. Pages that are useful to one user will be useful to other users who have the same interests.

In what follows, we focus on the evaluation procedure in the social filtering system. The basic mechanisms can be classified into those that are conscious and those that are unconscious. These methods have some problems, however cognitive overhead on users is increased in conscious evaluation, and it can't obtain user interests precisely using only the number of accesses in the unconscious way.

3. Extracting User Interests

In this section, we focus on how to evaluate for accessed pages automatically. To get this information without imposing a cognitive overhead, and successfully extracting the evaluation, we think it is important to make use of following information.

- Browsing history
  Users inside a fire wall can't access the outside network, so they need to access such networks by using proxy servers. These servers store user access logs as to what server they visit when. Using these logs, we extract a browsing history for each user.
- Event history
  This records the user's input history during WWW access using the browser, and it can record every action , i.e., mouse button press and release, and mouse motion, with a time record.
- Recognition history
  When users leave certain pages open for a long time, it is natural to assume they are interested in those pages. But it isn't good to evaluate pages purely on the basis of their access time. We use image recognition to decide whether users face the screen or not. This face-direction information specifies the real viewing time for each user.

For getting this information, we made a prototype using a mechanism which records events in the X-windows system and the results of face recognition. This system analyses the users actions by referring their browsing histories, making use of the above information (Figure 1). In the following subsections, we describe all parts in detail.

Figure 1 : System architecture

3.1 Browsing History

Our proxy server obtains the user's browsing history, that is, where they visit and how long they stay, and it stores the data. In general, user actions can be understood using this information, browsing details aren't clear using only these logs, because they don't include personal identification, so it is impossible to identify who is accessing the target information. This proxy server consists of clients and server systems which is written in Java. These clients receive requests which users using their browser make to access outside information, and send these requests to the real information server. In this session, they transfer session histories to the mediate server system that is written in HORB [Horb 97], which is equipped with the ORB (Object Request Broker) mechanism. Using this mechanism, it is easy to organize the communication between clients and server.

```
[Proxy Log]
 Start-Time Refer-Time  ID URL
 13:21:55  00046     1 http://www.foo.com/renewal.html
 13:22:41  00015     2 http://www.foo.co.jp/
 13:22:56  00036     3 http://www.foo.or.jp/music/
 13:23:32  00335     4 http://www.foo.ac.jp/
 13:25:47  00070     5 http://www.foo.org/
```

3.2 Event History

To follow user actions, it is advantageous to obtain a history of their activities, keyboard or mouse actions. We can't identify user events using only logs from the proxy server. We grab as X events by creating and putting a new transparent window on the browser, which receives user operations and sends them to the browser. It stores what events occur when these events occur. It stores the following information (Figure 2).

- event-name
  This makes clear what event occurs, for example, key press, mouse click and release, motion event, and so on.
- event-window
  This shows where the target event occurs. Users give actions to widgets which are parts of the target window. This information is given as the widget name.
- time-stamp
  This clarifies what time and for how long the target event occurs.

With this information, we can resolve whether users are interested in pages or not. We investigated user operations during browsing, and attention to the special operations and information is effective for extracting user interest, i.e., saving file, adding to the bookmark list, and having the page visit for a long time.

```
[X Event Log]
 Event-Name Event-Window Time-Stamp
 ButtonPress Text-Area    514631420
 ButtonPress File        514634090
 ButtonPress Text-Area    514636600
 ButtonPress Bookmarks    514640250
 ButtonPress Text-Area    514642720
```

Figure 2 : Extraction of X event

3.3 Recognition History

Use of browser logs and event histories is insufficient for identification, because there are some exceptions when deciding in what pages the user is interested. It is natural to assume that users are interested in WWW pages at which they stay for a certain time. But we can't judge how long user's attention remains directed at the page only from proxy logs and event histories. For example, it is considered that users remain when there are no keyboard or mouse events, but in fact they may be out of their seats, reading a book, etc. Thus we need more information to decide the real interest time, namely, to making use of the direction in which the user

faces. Facing the screen and remaining to stay in the browser window can be regarded as staying at the page and as interest in this page. We make use of image recognition to identify the direction of the user's face. Figure 3 shows a view of the face recognition process. The upper-left image is the original face image, and the lower-right image is the binary face image. In this binary image, some points are emphasized, that is, eyes and mouth. If users face the screen, the relative locations of eyes and mouth is fit a pattern that is similar for everybody. The following data shows a sample recognition log. A user can be considered to face the screen if there exists the string "eye_loc" in the evaluation.

[Face Recognition Log]

```
Event-Time Face-Location          Evaluation
13:22:10  face_loc(90, 63, 216, 189) eye_loc(111, 149, 156, 191)
13:22:10  face_loc(90, 63, 216, 189) eye_loc(111, 150, 156, 190)
13:22:10  face_loc(90, 63, 216, 189) out_of_mouth
13:22:11  face_loc(90, 63, 216, 189) eye_loc(112, 149, 155, 190)
13:22:11  face_loc(90, 63, 216, 189) eye_loc(111, 149, 156, 191)
```

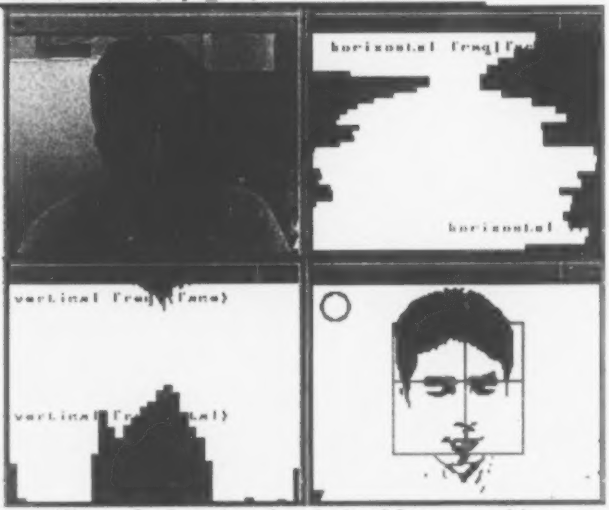


Figure 3 : Samples of analysis of face recognition

3.4 Action Identifying and Estimation

Using the above information, user interests are identified in the following way.

Step1. For each user, it extracts the relationship between user access time and degree of user interest evaluated consciously. For example, if a user refers to a page for long time, it is evaluated as high degree by the user.

Step2. It stores the X event log and proxy server for each user, and then it extracts access logs showing where each user browses.

Step3. Using the result of face recognition, it calculates real attention time for each page, that is, how long in fact users view the target pages.

Step4. It extracts URLs (Uniform Resource Locator) from access log which users add to the bookmark-list or save the disk. These pages can be considered to have the high interest for them. It marks these as special pages for the corresponding user.

BEST COPY AVAILABLE

Step5. It calculates the page evaluations using the correlation between real viewing time and the degree of user interest. For example, this user has high interest the URL in the number 4, because he accesses this page for 133 seconds, which is enough time to assume that he has a high interest, according to the statistical analysis between user interests and access time.

[Modified Proxy Log]

```
Start-Time Refer-Time  ID URL
13:21:55 00044      1 http://www.foo.com/renewal.html
13:22:41 00013      2 http://www.foo.co.jp/
13:23:32 00133      4 http://www.foo.ac.jp/
13:25:47 00068      5 http://www.foo.org/
```

4. Evaluation

In this section, we discuss the performance of our action identification and page evaluation. This performance check consists of comparing how well the system chose those pages preferred by users, with our technique and without it. This comparison is measured using the precision, that is, dividing the number of pages which match between the user and the system evaluation by the number of pages which users access. The following tables represent the precision between system results and user evaluations.

Result without revising access time/Result with revised access time

USER EVAL / SYSTEM EVAL	1	2	3	4	5
1	58/66	4/6	0/2	0/0	0/0
2	26/18	23/22	6/6	2/2	0/0
3	0/0	3/2	1/4	0/0	0/0
4	0/0	0/0	6/1	4/5	0/0
5	0/0	0/0	0/0	1/0	2/2

[Precision : 64.6 %/76.2 %]

Table 1 : Evaluation

These tables shows the relationship between the result without revising the access time and the result with. These tables clarify that it is efficient to use face recognition, since it increases the precision of the matching rate by about 11.6%. Therefore, our method is useful for identifying user interests.

5. Discussion

There are several methods for realizing a social filtering system [Resnick 94]. For example, users can judge whether pages are valuable or not, or the value of pages can be judged from the number of times they are referred to. The problem of the former method is the high cost for users, to judge every page, and the problem of the latter method is that it is not clear whether target pages have real value for users by judging only the number of times. Our method automatically calculates the value for pages according to the real viewing time using non-verbal information such as the X event log and face recognition log, so it can obtain true user interests without extra cost, and it is a way for users to be able to have their interests extracted unconsciously.

6. Conclusion

This paper describes a basic investigation into achieving a social filtering system, and we realized how user interests can be extracted by reference to WWW information. For building up the social filtering system, this interest extraction let us identify the user's intent automatically. It is also effective to make use of non-verbal information such as keyboard, mouse, and face direction for the users actions. In the future, there will be other non-verbal information that specifies user actions even more accurately, so it is necessary to investigate them next.

References

[firefly 96] firefly (1996). http://www'.firefly.com/

[Horb 97] S. Hirano (1997). HORB: Distributed Execution of Java Programs. *Worldwide Computing and Its Applications*, Information Processing Society of Japan (IPSJ), Tsukuba, Japan, 29-42.

[Malone 87] T. Malone, K. Grant, F. Turbak, S. Brobst, & M. Cohen (1987). Intelligent Information-Sharing Systems, *Communication of the ACM*, 30 (5), 390-402

[Resnick 94] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, & J. Riedl (1994). GroupLens: An Open Architecture for Collaborative Filtering of Netnews. *Computer Supported Cooperative Work*, ACM, Chapel Hill, North Carolina, USA, 175-186.

[Shardanand 95] U. Shardanand & P. Maes (1995). Social Information Filtering : Algorithms for Automating "Word of Mouth". *Conference on Human Factors in Computing Systems*, ACM, Denver, Colorado, USA, 210-217

[Susaki 96] S. SUSAKI, G. Kikui, & Y. Hayashi (1996). Navigation Interface in Cross-Lingual WWW Search Engine, TITAN. *AUUG & Asia-Pacific World Wide Web Conference*, Charles Sturt Univ., Melbourne, Australia, 352-356.

The Online Learning Academy

Suzanne Liebowitz Taylor, Donald P. McKay
Lockheed Martin C2 Integration Systems
Frazer, Pennsylvania, USA
{Suzanne.L.Taylor, Donald.P.McKay}@lmco.com
Ann Culp
Educational Technologies
El Segundo, California, USA
arc@interworld.net
Stephen Baumann, Karen Elinich
The Franklin Institute Science Museum
Philadelphia, PA USA
{baumann,kelinich}@fi.edu

Abstract: This paper describes the Online Learning Academy (OLLA), a WWW-based presence which supports the use of telecomputing in the classroom. Initial results from pilot use with twenty elementary schools teachers within the Department of Defense Educational Activity (DoDEA) schools during the 1996-1997 school year are presented.

1. Introduction

The proliferation of computer technology and Internet connectivity in K-12 schools creates a wonderful opportunity to connect educators and students to each other and to real world learning experiences, investigations, and explorations. However, the WWW is also an unfriendly, impersonal and often haphazard environment. It has little knowledge about a user or specific goals, lacks consistent organization of resources, and has little quality control. Teachers who lack technical sophistication and have goals that do not translate well to WWW search queries are at a disadvantage. If unable to find the appropriate information, they are likely to abandon the resource—and miss the opportunity to enrich their student's learning experience through technology.

The Computer Aided Education and Training Initiative (CAETI) [CAETI 1997] under Defense Advanced Research Project Agency (DARPA) sponsorship supports the advancement of computer technology for effective education and training. The Department of Defense Educational Activity (DoDEA) K-12 schools in four school complexes in Europe were pilot sites for the educational initiative. DoDEA serves the DoD military and civilian dependents' educational needs from preschool through high school in the United States and overseas. The DoD Dependent Schools (DoDDS), the overseas component of DoDEA, are similar to U.S. school systems in terms of student population and demographics. However, DoDEA has one significant difference from the U.S. school systems -- the 172 DoDEA schools are geographically dispersed throughout fourteen countries.

The potential for DoDEA to use the WWW to meet its challenges is great. The Internet may be used effectively to support interaction among the DoDDS teachers and students as well as to access current information and to stay abreast of technology. However, simply putting computers in classrooms, wiring a school and providing an Internet connection is not sufficient. Effective use of this technology will occur when the teachers understand how to integrate it into everyday practice and *want* to use it. Acceptance of technology in the classroom will be achieved when it is both relevant to educational goals and comfortable to use.

Lockheed Martin, Educational Technologies and The Franklin Institute Science Museum have developed the Online Learning Academy (OLLA) [OLLA 1997a, OLLA 1997b] as part of the CAETI program. OLLA is a WWW environment which supports the effective use of telecomputing and the Internet in the classroom:

by connecting teachers to each other and Internet educational resources,
by fostering the use of online resources and collaboration to enhance the classroom experience,
by encouraging and enabling the sharing of classroom experiences, and
by supporting and mentoring educators for all the above goals.

To engage the teachers, OLLA featured several online, thematic educational resources which provided the pilot teachers with many ideas on how to incorporate online resources into their classroom instruction. The remainder of this paper describes the Online Learning Academy and its use in the DoDEA testbed.

2. Deploying WWW Technology

Deploying WWW-based technology for effective use in the classroom requires a number of critical components. One is exploiting the client and distributed server architecture supported via the WWW. Technically, delivery of WWW content into the classroom is simple a web browser suffices. However, teachers and students rightfully need to view this interface as their portal into the wide open spaces of the Internet. As such, the WWW client is viewed more as their *virtual point of contact* or *launch point* onto the WWW than as a web browser. The importance of this observation is that the client needs to be an analogue of the place they are, namely their individual classroom and school. The more that the delivery mechanism is organized and tailored towards the school environment, the more effective and relevant the content delivered via the client will be.

The inherent distributed architecture of the WWW can be exploited in two key ways. First, for providing access to many rich and relevant educational sites, and, second, for providing a flexible and scaleable deployment into schools. General WWW resources are not typically well-constructed for educational use. For example, the need to support websites via advertising is a potential distraction to a teacher or a student. The delivery of educational resources needs to be mediated by providing server-based sites which function as well-founded and educationally relevant points of collaboration. Well-organized collections of topic-oriented general resources are also a mechanism for supporting educational use of WWW resources. The WWW as an infrastructure is inherently flexible as URLs can reference local or remote resources. The key is that the delivery to the teacher and the classroom is robust while the school infrastructure evolves, and, given the rapid and continual advance in network and computer technology, the ability to adapt the system over time is crucial.

3. The Online Learning Academy

OLLA is a virtual presence in the classroom which serves as a portal to the Internet. OLLA is a WWW Intranet environment which helps educators find relevant educational resources quickly, incorporate them easily into their daily classroom activities, and publish and share the results of these activities with others. Through a partnership among the application developers, the educational technologists, the curriculum specialists (in our case, science) and the end users (teachers), the success of OLLA project is based on the deployment of its three important components:

appropriate content - collections of organized educationally relevant resources, collaborative/targeted activities and user-contributed material,
continual professional development - a combination of on-site formal and informal sessions and continual online support, mentoring and presence, and
technology which supports these goals almost seamlessly quickly becoming natural to the user.

OLLA users find the graphical interface, which is organized with customized classrooms and a resource center, a familiar environment that is easy to use. As the teachers begin to integrate OLLA technology into their classrooms, they are starting to look differently at the way they teach. Both teachers and students find a great deal of information to supplement their textbooks, and consequently expand their knowledge base beyond what was possible in the past. OLLA also encourages users to become producers of Internet information

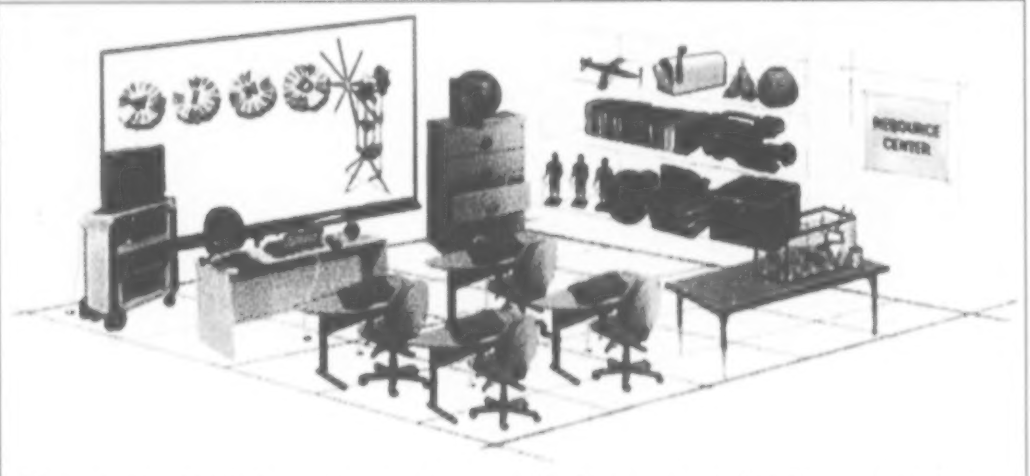


Figure 1: The OLLA classroom serves as a personalized interface to all OLLA resources, including current projects, journals and mailing lists.

instead of just consumers. By publishing students' work on the WWW, we believe OLLA could become a motivational tool for students as they discover their accomplishments will be viewed by other students.

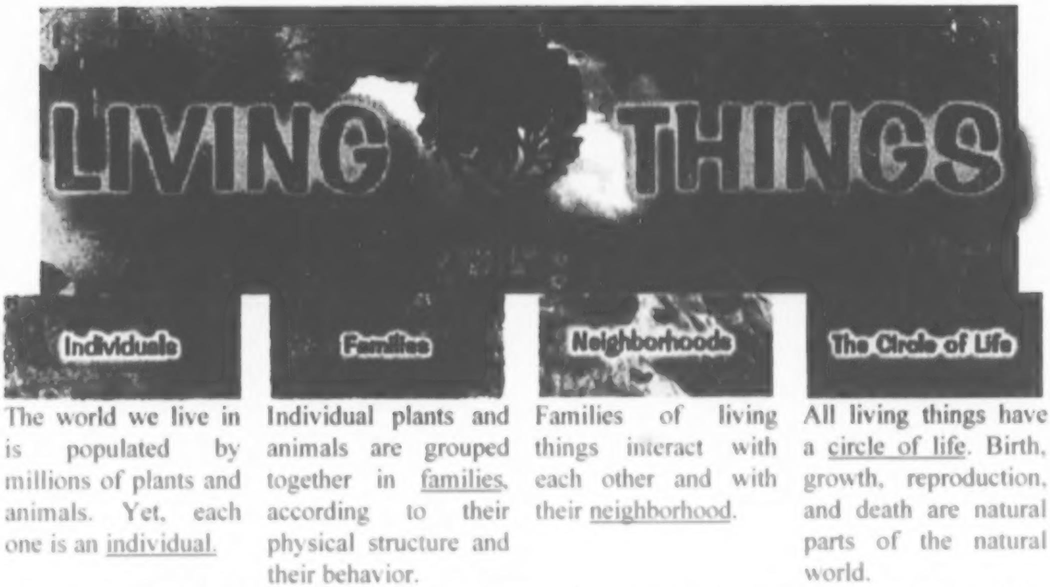
4. A Brief OLLA Tour

As teachers enter OLLA, they are prompted for a personalized logon and password. OLLA uses this information to take the teacher to a personalized virtual *Classroom* (Figure 1), and to support pilot usage data collection. Once there, the teacher uses OLLA in a variety of ways. By clicking on the file cabinet, the teacher views original lesson plans, complete with teacher-selected Internet links. There are also activities and additional plans written by other teachers which may offer new ideas to implement in their classrooms. Featured activities and units of study (Section 0) are displayed on a white board at the front of the classroom for quick and easy access to those resources. OLLA also provides a separate classroom interface for students which is shared by all students in the same class. The student classroom provides easy access to the *Kids Did This!* gallery and current projects. *Kids Did This!* is an organized collection of WWW publications and a favorite spot for viewing other students' work.

From the classroom, a teacher may click on the door which opens directly to the *Resource Center*. In the Resource Center, the teacher finds a wide array of Internet resources and teacher activities which have been organized by subject area. Teachers may go directly to the topic of interest, or query the Resource Center to find the relevant information they seek. Available from the Resource Center and the Classroom are links to current publications such as newspapers and periodicals, which allow teachers to bring up-to-date information into the classroom. Reference resources, such as an online dictionary, thesaurus, maps, World Fact Book, and Bartlett's Book of Quotations are only a click away. Using the mouse, teachers easily access several Internet search tools which allow them to find additional resources to supplement interests and activities.

Other valuable features of OLLA include the teachers' *mailing list* and *personal journals*. The mailing list allows teachers to communicate with other OLLA teachers. Teachers are encouraged to use this mailing list to share information and ideas and to solicit collaboration in classroom projects and activities. Journals are provided so teachers may write reflections and thoughts about classroom projects. While the journals are personal writings, they may be read by any OLLA teacher who wishes to learn from others' experiences with similar projects or studies. In addition, a *Problems* mailing list is linked to the headers and footers of every page so that technical problems can be quickly reported, tracked and addressed.

BEST COPY AVAILABLE



The world we live in is populated by millions of plants and animals. Yet, each one is an individual.

Individual plants and animals are grouped together in families, according to their physical structure and their behavior.

Families of living things interact with each other and with their neighborhood.

All living things have a circle of life. Birth, growth, reproduction, and death are natural parts of the natural world.

Figure 2: The *"Living Things"* unit includes selected and organized web resources, presented around the theme of ecosystems.

Each teacher creates a profile page where personal information, pictures and contact information are posted. These pages are located in the *Members List* and are a wonderful way for teachers to locate colleagues in different schools who teach the same grade levels or content. Teachers have also found this a useful place to link classroom pages, which often contain student portfolios. The *Passport* matchmaker system, an interface to a searchable database of educators, connects the DoDEA teachers with other (non-OLLA) stateside teachers through active searches on user profiles.

Two forms of search are available, depending where you are in OLLA. Through the headers and footers on nearly every page is a general search facility. On certain designated pages (such as *Help* and the *Resource Center*) localized searching can be triggered to seek information from pages associated with a particular topic or area.

5. Online Units of Study

There is no shortage of educational resources on the WWW. However, teachers may lack the time and skill to locate and evaluate them. OLLA units of study, like *"Living Things"* [TFI 1996] and *"Wind: Our Fierce Friend,"* [TFI 1997] include selected and organized web resources, presented around a theme. For example, in *"Living Things,"* (Figure 2) the theme of ecosystems is considered. Links to existing web resources are strategically placed within newly created content that facilitates hands-on classroom investigation of the theme. Plans for growing seeds in the classroom are complemented with links to online plant resources. Tips for raising fruit fly colonies are supported with links to fruit fly physiology resources.

OLLA thematic units enable teachers who may be novice technology users to incorporate online resources into their classroom instruction. At the same time, the units encourage approaches to hands-on classroom investigations. The availability of, and access to, organized online units of study may have a significant impact on the acceptance of new technology by veteran teachers. OLLA offers them easy access to instructional resources, convenient tools for communication with the online educational community, and rich opportunities and ideas for collaborating with schools around the world. Easy, convenient, and rich may be significant adjectives as teachers begin to articulate their future desires for technology in their classrooms.

BEST COPY AVAILABLE

6. OLLA Use in Pilot Project

OLLA was in pilot use during the 1996-1997 school year with about twenty elementary school teachers. The participating DoDDS school complexes include three sites in Germany and one in Italy. Initial use in fall 1996 included eight pilot teachers from the 2nd and 5th grades. In the spring of 1997, twelve 3rd and 4th grade teachers joined the project. The core OLLA components consist of a set of HTML content pages and Common Gateway Interface (CGI) programs which are accessible to the user via an HTTP server. The search component uses the Harvest indexing and retrieval engine [Harvey, Schwartz and Wessels 1994].

The Franklin Institute Science Museum facilitates and encourages hands-on, collaborative science instruction. Museum-developed units of study emphasize inquiry-based teaching and learning (Section 0). So far, OLLA has featured two such units: *"Wind: Our Fierce Friend"* and *"Living Things."* In both, the unit of study offers connections to online information, areas for communication, potential for collaboration, and places to share student and teacher work. There are deliberate differences between the two, however. *"Wind"* is based upon all classrooms receiving the same hands-on materials so that students undertook common activities and then used the online unit to communicate and share their results and experiences. In *"Living Things,"* teachers and students use their own existing classroom materials, such that students undertake completely different classroom activities. The online unit is the common element and a bridge for sharing their diverse perspectives on the theme. Additionally, *"Living Things"* offers a more overt connection between classroom activities, national standards and curricular themes, while *"Wind"* is directed toward completely open-ended investigation.

Professional development and support for the teachers participating in the OLLA project consists of on-site formal staff development sessions, informal follow-up visits to classrooms, accessible online documentation and help, and continual mentoring and assistance via e-mail. Formal staff development included the basic use of OLLA, instruction for some general technical skills and assistance in preparing a technology plan to

THE WIND

Wind can cause tornadoes. Wind
can be strong and weak. My
brother thinks the wind is a person.
I know it is not. Wind can cause
total destruction.

Figure 3: A variety of creations result from participation in the thematic activities.

BEST COPY AVAILABLE

integrate OLLA and the featured thematic units into classroom instruction. The teachers have progressed significantly in their understanding of technology and its relevance to their classroom goals. The classroom portfolios based on both units include poetry, prose, drawings and photographs of activities such as building pinwheels, constructing windmills and taking nature walks (Figure 3).

7. Conclusions and Future Work

The WWW OLLA implementation is readily scaleable in terms of number of teachers, adding new schools, classrooms and content. Our goal is to further enable teacher independence by coaching teachers into the role of mentors for newer participants, as the OLLA project continues to grow. Many of the pilot teachers have begun take a more proactive role as they become more comfortable with the technology.

Within the CAETI program, each pilot project fielded in a school underwent a formal evaluation. Initial results from OLLA use during the 1996-1997 school year indicate initial success. During the evaluation surveys, the teachers reported that:

OLLA provided the means for communicating with other classrooms and teachers located at a distance.

OLLA helped them to reach students who were difficult to reach using other approaches.

OLLA is a great motivator for students and teachers alike.

OLLA is changing the way teachers think and teach as a result of seeing other possibilities.

Technology adoption by teachers is extremely difficult if it is imposed and is not relevant to what the teacher needs and does in the classroom. After fall pilot useage, over 80% of the teachers in OLLA classrooms indicated they would use OLLA next year if it is available; the remaining 20% would probably use it—well above the normal 30 to 40% acceptance rate in the literature. OLLA's adoption indicators suggest the deployment and supported use of highly targeted and relevant technology via educational technology specialists is a highly effective model. Perhaps it can break through the technology adoption rate barrier in schools.

In addition, we have a companion research and development effort [Pastor, Taylor, McKay and McEntire, 1997] which complements our goals of enabling appropriate, timely, customized access to Internet resources through a set of intelligent resource agents which perform a variety of tasks related to supporting and enhancing the use of the Internet as an educational tool. The resulting system is accessible from within a WWW infrastructure and was successfully integrated with OLLA. Aspects of this technology will enhance OLLA as the project matures.

8. References

[CAETI 1997] The Computer Aided Education and Training Initiative. *http://www.dmso.mil/CAETI/*

[Hardy, Schwartz and Wessels 1994] Hardy, D.R., Schwartz M.F., & Wessels, D. (1996). *Harvest User's Manual*. Technical Report CU-CS-743-94, Department of Computer Science, University of Colorado, Boulder, Colorado, January 1996. *http://harvest.transarc.com/afs/transarc.com/public/trg/Harvest/user-manual/user-manual.html*.

[OLLA 1997a] Lockheed Martin C2 Integration Systems (1997). *Online Learning Academy (OLLA), An Introductory Tour*, 1997.

[OLLA 1997b] Lockheed Martin C2 Integration Systems *(1997). Online Learning Academy (OLLA), Users Guide*, 1997.

[Pastor, Taylor, McKay and McEntire, 1997] Pastor, J., Taylor, S., McKay, D., & McEntire, R. (1997). An Architecture for Intelligent Resource Agents. *Proceedings of Cooperative Information Processing Systems Conference (COOPIS - 97)*, Kiawah Island, South Carolina.

[TFI 1996] The Franklin Institute Science Museum (1996). *Wind: Our Fierce Friend*. *http://www.fi.edu/tfi/units/wind/*.

[TFI 1997] The Franklin Institutee Science Museum *(1997) Living Things* http://www.fi.edu/tfi/units/life/.

Acknowledgments

Success of this project is largely due to the enthusiastic participation of the OLLA pilot teachers in Hanau, Würzburg and Kaiserslautern, Germany and Aviano, Italy DoDDS schools. The authors would like to acknowledge the invaluable contributions to this project from Roslyn Nilson, Christian Polizzi, Christine Baker, Jon Pastor, Robin McEntire, Peter Stevens and Peg Duffy at Lockheed Martin and Kelly Knight at The Franklin Institute Science Museum. This work was funded in part by the Defense Advanced Research Projects Agency under contract N66001-95-8631.

On the Use of Librarians Selection Routines in Web Search.

Avgoustos. A. Tsinakos

Kostantinos. G. Margaritis

Department of Informatics
University of Macedonia
54006 Thessaloniki
GREECE
Tel : +30-31- 891 891
E-mail: tsinakos/kmarg@kirki.it.uom.gr

Abstract: Information retrieval on the Web has a major obstacle: although data is abundant, it is unlabeled and randomly indexed. This paper discusses the implementation of a consultative Web search engine that minimizes the expertise level that is required from a user in order the latter to accomplish an advance search session. The system takes advantage of the meta-knowledge (Selection Routine), used by expert librarian searchers and apply it to a heterogeneous search space such as CD-ROM Data Bases and WWW based environments acting as an intermediary expert system.

Introduction

Due to the rapid growth of data, taming the information relaying on the Internet repositories has become a difficult and time consuming process. Beyond the Web based search engines such as Lycos and AltaVista that have appeared on the net, some other sophisticated mechanisms have been developed, to confront with the problem of information retrieval. ALIWEB (Archie-Like Indexing in the WEB) [Koster 1994], GENVL (an interactive hierarchical system for cataloguing Web resources in a sense of "Virtual Libraries") WWW Worm - (a resource location tool) [McBryan, 1994] are some representative examples.

Recently several new software products have emerged on the Web space. The common target of those products is to reduce the user effort during a information retrieval session on one hand, and on the other, to increase the productivity and accuracy of the retrieval process using AI and parallel searching techniques. Intelligent Agents used by the MORE LIKE THIS [MORE LIKE THIS] and AUTONOMY [AUTONOMY] products, are trained by the user and released in web space in order to locate an derive the requested term-concept. Additionally meta-search engines such as WEB COMPASS 2.O [WEB COMPASS], MetaCrawler [MetaCrawler] and ECHO SEARCH [ECHO SEARCH], are applying parallel searching on pre-selected Web based search engines and filter the retrieval set by eliminating the duplicates. Web miners are another category of information retrieval systems that relay on a combination of test queries and domain specific knowledge to automatically learn descriptions of Web services such as product catalogues or personal directories. Internet Learning Agent (learns to extract information from unfamiliar resources by queering them with familiar objects) [Perkowitz & Etzioni, 1995] and Shopbot (learns to extract product information from Web vendors) [Doorenbos et al, 1996] are such systems. Internet Softbot can automatically extract information or learned descriptions collected by such intelligent agents [Etzioni, 1994]. Some of the latest products in knowledge-based information retrieval technology for the WWW are: FAQFinder which is an automated question-answering system that uses the FAQ files which are associated with many USENET newsgroups, in combination with FindMe and RentMe systems (market search agents) [Bruke & Hammond et al 1997]. The increasing use of AI techniques in the information retrieval process, reveal a new tendency and need for more intelligent and flexible systems with a high degree of search expertise regarding the procedural and declarative knowledge, in order to perform a search task.

Similar requirements have been outlined in the area of database and on-line search, by the catalogers and reference librarians. There is a number of inherent characteristics of on-line catalogs that make them difficult to use, especially when someone was seeking subject information. [Bates, 1972] [Borgman 1986] [Connel, 1991]. The identification and characterization of the knowledge used by experienced librarians during a subject searching process in on-line catalogs, is considered an important topic for investigation since an understanding of the specialized knowledge used by the librarians may facilitate the design of more usable systems [Connel 1995]. Tackling this problem, a number of systems had been implemented like Source Finder [Bailey 1992] and Reference Expert [Myers 1994].

In this paper an effort to combine the needs of both librarian reference search and Internet information retrieval, is attempted. The main idea of the system that will be discussed is its intelligence of taking advantage of the meta-knowledge called "Selection Routine", used by expert librarian searchers, in order to construct a search plan. Furthermore this plan will be applied on heterogeneous search spaces such as Data Bases and WWW based environments. An analysis of the rules which consist the Selection Routine and system's architecture for the co-operation of system's core and the retrieval mechanisms follows.

Selection Routine

The intellectual components of a typical on-line search can be analised in to three basic stages. I) The definition of query structure stage II) The selection of search keys stage and III) The feedback review stage [Fig 1].

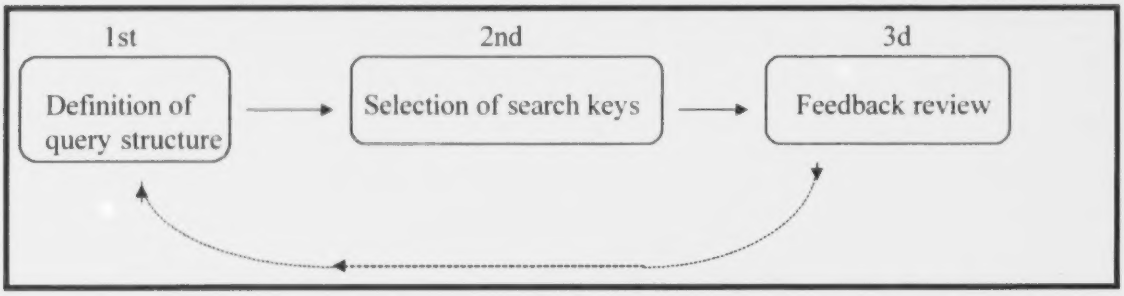
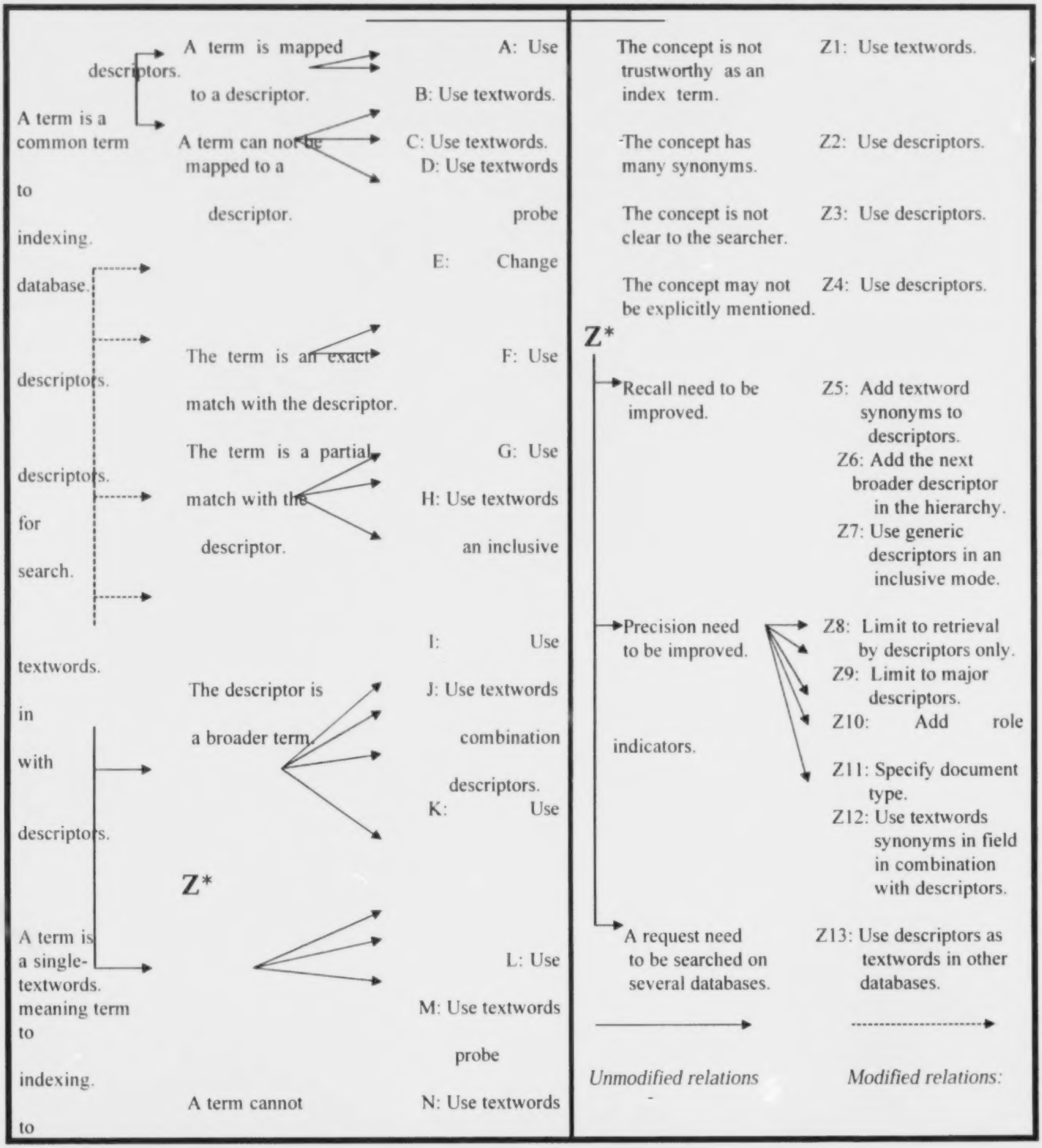


Figure 1
Components of on-line search

At the second stage, the search expert having clarified both the Semantic and the Pragmatic aspect of the request, proceeds to the construction of the search strategy. Four layers comprises the search strategy plan. At the first layer the expert selects the Data Base to be searched towards user's request. At the second layer expert considers regarding which terms- search keys will be used during the search process. At the third layer expert opine which search key will be entered first, and finally at the forth stage , which at the same time affects the third stage of Feedback review, experts mediate how to review unsatisfactory results.

The basic dilemma of the librarian search experts, is the appropriate use of free text search or the controlled vocabulary search according the type of the search key. If a search key is a single meaning term (uniquely defined and specific to the concept that represent), then using free text search seems to be the most promising choice to be followed. On the contrary in the case where the search key is a common term having a broad and vague meaning, then free text search destroys the relativeness and preciseness of the retrieval set, and thus controlled vocabulary search is preferred. The advantage of the controlled vocabulary search type, is the use of descriptors which are single meaning terms used for thesaurus construction in databases. Many concepts are accurately indexed under such descriptors. Therefore a crucial point for the performance of the search process, is the selection of the search key that will be used. On the research project *"Searchers' Selection of Search Keys* (part I II III)", of Raya Fidel, is expressed the idea that expert searchers usually follows some general rules in order to conclude which search key to select before the search session is initialised. This set of rules is defined as *Selection Routine* [Fidel 1991]. An overview of the decision tree of the Selection Routine and the corresponding paths from the initial assumption to the final decision is sown in Fig. 2.

As an explanatory example of the searchers selection routine, consider the case where a search term is a common term and is mapped to a descriptor. Then this descriptor will be used as a search key instead of the common term (Case A).This fact implies a controlled vocabulary search type. However when a common term can not be mapped to a descriptor then fact implies the use of free text search (Case C). In the above figures the continuous lines represent the relations of the initial assumption to the final decision as they are shown in the original paper, while the dotted lines correspond to the modified relations that are .


A term is mapped
A: Use
descriptors.
to a descriptor.
B: Use textwords.
A term is a
common term
A term can not be
C: Use textwords.
mapped to a
D: Use textwords
to
descriptor.
probe
indexing.
E: Change
database.
The term is an exact
F: Use
descriptors.
match with the descriptor.
The term is a partial
G: Use
descriptors.
match with the
H: Use textwords
for
descriptor.
an inclusive
search.
I: Use
textwords.
The descriptor is
J: Use textwords
in
a broader term.
combination
with
descriptors.
K: Use
descriptors.
Z*
A term is
a single-
L: Use
textwords.
meaning term
M: Use textwords
to
probe
indexing.
A term cannot
N: Use textwords
to
The concept is not
trustworthy as an
index term.
Z1: Use textwords.
The concept has
many synonyms.
Z2: Use descriptors.
The concept is not
clear to the searcher.
Z3: Use descriptors.
The concept may not
be explicitly mentioned.
Z4: Use descriptors.
Z*
Recall need to be
improved.
Z5: Add textword
synonyms to
descriptors.
Z6: Add the next
broader descriptor
in the hierarchy.
Z7: Use generic
descriptors in an
inclusive mode.
Precision need
to be improved.
Z8: Limit to retrieval
by descriptors only.
Z9: Limit to major
descriptors.
Z10: Add role
indicators.
Z11: Specify document
type.
Z12: Use textwords
synonyms in field
in combination
with descriptors.
A request need
to be searched on
several databases.
Z13: Use descriptors as
textwords in other
databases.
Unmodified relations
Modified relations:

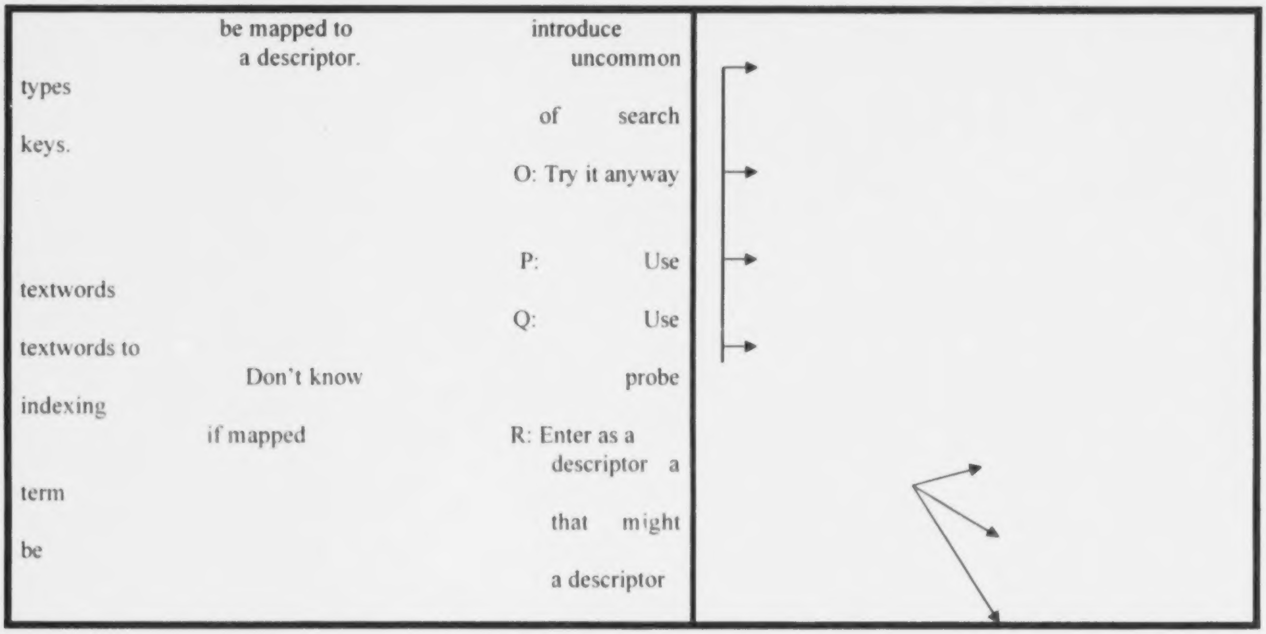

be mapped to
a descriptor.
introduce
uncommon
types
of search
keys.
O: Try it anyway
P: Use
textwords
Q: Use
textwords to
Don't know
probe
indexing
if mapped
R: Enter as a
descriptor a
term
that might
be
a descriptor

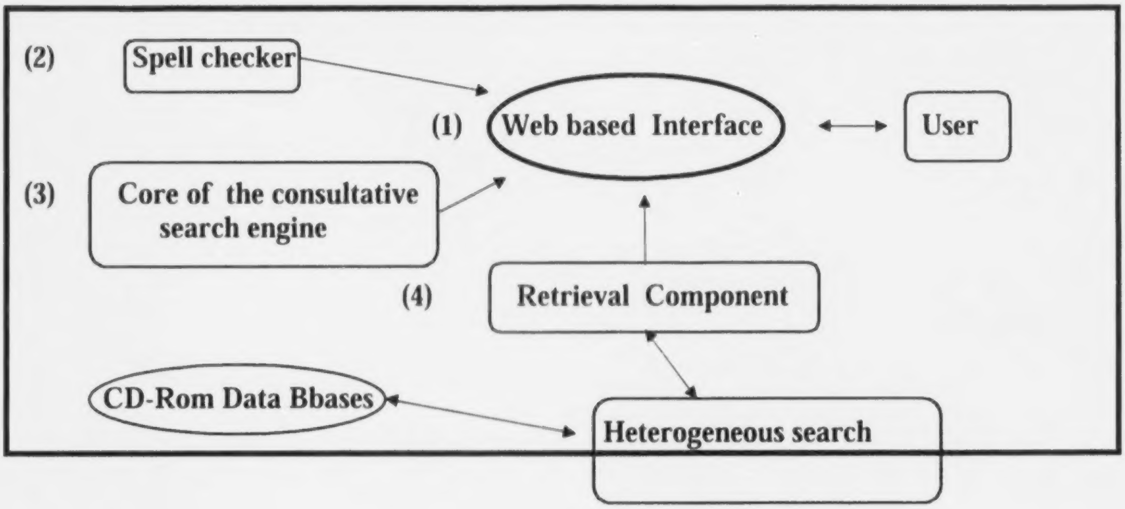
Figure 2
Decision tree of the Selection Routine

System description.

The system comprises of the following components [Fig 3].

1. Web based Interface: This component is the front end user interface where the user can interact with the system and define the desirable search term. Additionally user is interviewed by the system in order the semantic and the pragmatic aspect of the search to be clarified.
2. Spell checker: A spell checker is used in order to eliminate misspelled search terms. Speller fires optionally after user's suggestion.
3. Consultative core: This component includes the knowledge base of the system (Selection Routine, Metaknowledge rules), and interacts with the retrieval component.
4. Retrieval Component: Retrieval component combines a variety of retrieval tools which co-operate with the consultative core in order the retrieval set to be achieved.

A further description of the Consultative core component will follow.


(2)
Spell checker
(1)
Web based Interface
User
(3)
Core of the consultative search engine
(4)
Retrieval Component
CD-Rom Data Bbases
Heterogeneous search

Internet repositories
environments

Figure 3

System description

In a typical search session user accesses the system via WWW. The system in order to clarify the Semantic and the Pragmatic aspect of user invoked search session, initialize an interview session with the user by displaying a number of form based questions. Filling these forms, user among others, is requested to define the search term, the repositories that he/she prefers to be searched (CD-Rom Data Bases or Internet or both), the corresponded topic to the search term (i.e. music or education) or if an expert's suggestion regarding the selection of an appropriate keyword is required (use of thesaurus). The defined search term can optionally be spell checked. Furthermore, consultative component using the selection routine, opines regarding the appropriated search key that will be used during the search session and the corresponding repositories that they will be reached. The results of the retrieval set are displayed to the user and in case that results are unsatisfactory, the whole search session can be refined.

Consultative core includes the knowledge base of the system in a form of *if... then...* rules and using is a forward chain inference engine constructs the search strategy. Two kinds of rule sets are included in this component. The first rule set represents the decision tree of the Selection Routine. This rule set affects the selection of the search key and the search type (free text or controlled vocabulary). The second rule set is the Metaknowledge rule set which have an effect on conflict resolution cases. A representative example of both rule sets is given.

Selection Routine Rule Set: In this example, cases A and B of Fig 2 are represented where the user defined term [term] is common term [CTR] and is mapped to a descriptor, so search expert can either use descriptors. [DSRC] to apply a control vocabulary search method (case A), or can use textwords [TXTWRD] to apply a free text search. So Rule_A corresponds to case A, and Rule_B corresponds to case B.

```
Rule_A: If
            is_CTR <term>
          & is_mapped_to_DSRC <term>
        then
            use_DSRC
```

```
Rule_B: If
            is_CTR <term>
          & is_mapped_to_DSRC <term>
        then
            use_ TXTWRD
```

Metaknowledge Rule Set: As it is earlier stated this rule set concerns conflict resolution cases. Conflict resolution in general, corresponds to the system "making up its mind" which rule to fire [Jackson]. During the contraction of the search strategy, it is very often the case where two or more rules are eligible to fire. In such cases meta-rules take effect in order to solve the conflict session by suggesting to the system which rule to fire first. Perceiving the above rules statements, Rule_A (RA) and Rule_B (RB), typical example of a conflict session can be noticed. Both left hand side premises of Rule_A (RA) and Rule_B (RB) are the same: *The defined term [term] is common term [CTR] and is mapped to a descriptor [DSRC],* while the right hand side premises are completely different. So it is obvious that in a forward chain session where these premises are true [T], both rules RA and RB will be loaded on the working memory [WM]of the inference engine and will be both eligible to fire causing a conflict to the system. At this point, meta-rules becomes activated. An example of the structure of meta-rules are the MR_ 1 and MR_11 rules

```
MR_1: If
            is_not_nil <WM>
          & RA and RB member_of <WM>
            not_need_improve_recall <T>
        then
            use_RA
        just
            Searcher almost always prefer to
            enter descriptor as search key in
```

```
MR_11: If
            is_not_nil <WM>
          & RA and RB member_of <WM>
            not_need_improve_recall <F>
        then
            use_RB
        just
            When the recall set using
descriptors
```

order to restrict the retrieval set.	*is poor, searchers prefer to use text-word search alternatively to increase the number of recalls.*

In this example MR_1 MetaRule, on the left hand side examines *if the Working Memory of the system is not empty*, in other words system has started the chaining, *if RA and RB are loaded on the WM*, causing a conflict, and additionally examines *if the retrieval set do not need improvement*, in order to assure that the retrieval set had not been obtained yet-search session is still on progress. In that case meta-rule MR_1 loans priority to RA in order to fire first and additionally provides to the user, the justification for this selection (optionally). In case where the retrieval set had already been obtained and considered to be poor or irrelevant, then MR_1 loans priority to RB to fire in order recalls to be improved (refinement session). Again the justification for this selection is available to the user.

Implementation issues

System implementation issues have been also addressed in the discussion of Mentor system [Tsinakos & Margaritis 1996]. Consultative core component of the system is being implemented in Alegro Lisp, while the front end interface of the system is hosted on a CL-HTTP server CL-HTTP is a full-featured server for the Internet Hypertext Transfer Protocol, implemented in Common LISP in order to facilitate exploratory programming in the interactive hypermedia domain and to provide access to complex research programs, particularly artificial intelligence systems [Mallery, 1994].

The Retrieval Component, in co-operation with the Consultative core component, reaches the appropriate repositories in order the retrieval set to be achieved. In case where the search session regards information retrieval from a database- CD-ROM, retrieval component uses the SilverPlatter information retrieval system for the Internet environment called WebSPIRS [WebSPIRS]. WebSPIRS provides potential to the user to search a remote CD-ROM database using WWW interface. In case where the retrieval is applied on Internet repositories, retrieval component uses a number of intelligent meta-search engines such as Quarterdeck WebCompass 2.0. Such meta-search engines can "work" in conjunction with popular search engines such as AltaVista, Yahoo, WebCrawler, Excite and Lycos, as well as many others and are able to filter summarize and categorize the acquired information. Retrieval Component can apply a search in both environments (CD-ROM Data Base and Internet), using at the same time the WebSPIRS and meta-search engines retrieval tools.

The ability of remote search of a CD-ROM database using WebSPIRS software, has been accomplished by using SilverPlatter's Electronic Reference Library Technology. Electronic Reference Library is a multi-user application server implementation of SilverPlatter's CORE technology. ERL client/server model provides local and wide area networking access to all SilverPlatter databases and enables easy loading of pre-indexed and ready to search information from CD-ROM or tape.

References.

1. [AUTONOMY] http://www.agentware.com
2. [Bailey 1992] Charles W. Bailey "The Inelligent Reference Information System Project" Information Technology and Libraries 11, September 1992, 237-44.
3. [Bates, 1972] M. J Bates, *"Factors affecting subject catalog search success"*, Unpublished doctoral dissertation, University of California, Berkely 1972, USA.
4. [Borgman 1986] C. L Borgman, *"Why are online catalogs hard to use? Lessons learned from information retrieval studies"*, Journal of the American Society for Information Science, 1986, 37, 387-400.
5. [Bruke & Hammond et al 1997] Robin Bruke and Kristian Hammond and Benjamin Young and Julia Kozlovsky, *"Intelligent Web Search Engines"*, PC AI Jan/Feb. 1997, 39-42.
6. [Connel, 1991]. Tschera Harkness Connell, *" Subject searching in on line catalogs: An explanatory study of knowledge used"*, Unpublished doctoral dissertation, University of Illinois, Urbana-Champaign 1991, USA.
7. [Connel 1995] Tschera Harkness Connell, *" Subject searching in on line catalogs:Metaknowledge Used by Experienced Searchers"*, Journal of the American Society for Information Science, 1995, 46 (7), 506-18.
8. [Doorenbos et al,1996] R.B. Doorenbos , O.Etzioni and D.S. Weld. *"Ascalable comparsion -shopping agent for the world-wide web"*, Technical Report, Dept. of Computing Science and Engineering, Univ. of Washington, Jan 1996 USA.

9. [ECHO SEARCH] http://www.iconovex.com/ECHO/ECHOS.HTM
10. [Etzioni, 1994]. O.Etzioni and D.S. Weld. *"Asoftbot-based interface to the Internet"*, Comm of ACM 37 (7), Jul. 1994, 72-76. http://www.cs.washington.edu /research/softbots.
11. [Fidel 1991]. Raya Fidel, *"Searchers' Selection of Search Keys: I. The Selection Routine, II. Controlled Vocabulary or Free-Text Searching, III. Searching Styles"*, Journal of the American Society for Information Science, 1991, 42 (7), 490-527.
12. [Jackson].Peter Jackson, *"Introduction to Expert Systems"*, Second Edition, Ch 8 pg. 142-143.
13. [Koster 1994] Martijn Koster *"ALIWEB - Archie-Like Indexing in the WEB"*, First International WWW Conference, May 1994, CERN, Geneva Switzerland http://www.nexor.co.uk/public/aliweb/aliweb.html
14. [Mallery, 1994] John C. Mallery *"A Common LISP Hypermedia Server"*, First International WWW Conference, May 1994, CERN, Geneva Switzerland http://www.ai.mit.edu/projects/iiip/doc/cl-http/server.html
15. [McBryan, 1994] Oliver McBryan, *"GENVL and WWWW: Tools for Taming the Web"*, First International WWW Conference, May 1994, CERN, Geneva Switzerland http://wwwmbb.cs.colorado.edu/~mcbryan/bb/summary.html
16. [MetaCrawler] http://www.metacrawler.com
17. [MORE LIKE THIS] http://www.morelikethis.com
18. [Myers 1994] Judy E. Mayers *"A newer version of Reference Expert, now beeing tested, has a keyword capability"*, Personal communication, Dec. 16, 1994.
19. [Perkowitz & Etzioni] M. Perkowitz and O. Etzioni. *"Category translation: Learningto understand information on the Internet"*, Fifteenth Internetional Joint Conference on AI, Aug 1995, Montreal CA, 930-936.
20. [Tsinakos & Margaritis 1996] Avgoustos A. Tsinakos and Konstantinos G. Margaritis, *"MENTOR Internet Search Advisor and Information Retrieval System"*, WebNet 96 - World Conference of the Web Society, Oct. 1996, San Francisco USA, 583-84. http://129.115.62.189:80/info/webnet96/html/114.htm
21. [WEB COMPASS 2.O] http://webcompass.qdeck.com
22. [WebSPIRS]. http://www.silverplatter.com/sampler/webspirs.html

Acknowledgments.

To Dr. Raya Fidel who is an Associate Professor at the Graduate School of Library and Information Science, University of Washington.

Modelling Alter-egos in Cyberspace Using a Work Flow Management Tool: Who Takes Care of Security and Privacy?

Reind P. van de Riet,
Hans F.M. Burg
Department of Mathematics and Computer Science
Vrije Universiteit, Amsterdam
e-mail: {vdriet, jfmburg}@cs.vu.nl

Abstract: Cyberspace will be inhabited by citizens, which when represented by objects are called Alter egos. These alter egos do business with institutions and shops. An important question is who takes care of Security and Privacy (S&P) rules.

In this paper we will assume that the processes these Alter egos are participating in are defined by WorkFlowManagement (WFM) diagrams. We will study how these WFM diagrams can be used to derive S&P rules in the form of authorization tuples. We will also sketch an architecture, using the Mokum system, in which one can prove that the S&P rules are kept. To translate this architecture in a global one involving secure communication in the Web is ultimately the goal of our research project.

1. Introduction

In a preceding paper [van de Riet&Burg96], presented at WebNet96, we have shown how Cyberspace can be considered as inhabited by alter egos, which are active objects representing people. In that paper we have shown how these alter egos can be modeled using COLOR-X [Burg96]. In particular we have discussed the question whether they can be held responsible. In a couple of other papers [van de Riet&Gudes96,van de Riet&Junk&Gudes97], we have dealt with the question of how to maintain S&P rules in an object-oriented system, like Mokum, which we use for demonstrations and simulations. In this paper we will show how Work Flow Management (WFM) diagrams (see e.g. [Georgakopoulos95]) can be used for specifying S&P rules and we will see how they can be maintained in Mokum programs generated automatically. We will briefly deal with the problem how in the global environment of the Web these S&P rules can be maintained.

WFM applications are already common in Business environment. We shall see in section 2 why WFM is also important to model the behaviour of alter-egos in Cyberspace, wherein we demonstrate that COLOR-X, developed in our group to model Information and Communication Systems using linguistic knowledge, can be considered as a WFM-tool, and we will see how S&P rules can be derived from COLOR-X diagrams. A typical example will be treated: filing and receiving payment for an insurance claim [Olivier96]. In this case the players may be at different locations: the submitter of the claim, the travel agent and the insurance company.

The next section gives a way how to look at alter-egos from an object point of view. In particular, an implementation will be discussed using the Mokum system with which it is possible to do simulations of the behaviour of alter-egos.

Other work on WFM and S&P can be found in [Atluri&Huang96] in which the authors use Petri net theory to represent security dependencies between tasks, defined in a WFM diagram (WFD), in order to derive and enforce multi-level security constraints.

2. Background on Workflow and Security

WFM tools are currently being used to specify how people and Information and Communication Systems (ICS or X) are cooperating within one organization. There are at least three reasons why WFM techniques are also useful in Cyberspace.

First, organizations tend to become multi-national and communication takes place in a global manner, also consultation of databases is done more and more globally; users have no idea where databases are located.

Second, more and more commerce is being done electronically. This implies that procedures have to be designed to specify the behaviour of the participants. These procedures may be somewhat different from ordinary WFM designs, where the emphasis is on carrying out certain tasks by the user-employees, while in commerce procedures are based on negotiating, promises, commitments and deliveries of goods and money. As we will see, in COLOR-X we have a notion: MUST, which is perfectly designed to represent these notions.

Third, people will be participants in all kinds of formalized procedures, such as tax paying or home banking. It is of greatest importance that matters around privacy protection are precisely defined. That means that the procedures have to be precisely defined also with respect to S&P.

This being said, how can we derive security and privacy rules from a WFD? In specifying tasks and actions of people working in an organization naturally involves also the specification of their responsibilities [van de Riet&Burg96a,b]. This is what WFDs usually do. Responsibility implies access to databases to perform certain actions on data of individuals. S&P rules come in two flavors: one is positive and concerns the access rights a certain user has or a group of users have according to some role. The negative flavor is the opposite: it comprises the rules which exclude users and groups of users from certain access rights. Translation of these principles to WFDs means: if a user or group of users has a certain responsibility to carry out a task involving information about certain individuals s/he has also the access right to the appropriate information.

In this paper we will deal with the process of an individual submitting a claim to an insurance company about a trip booked at a travel agent.

3. The Insurance-claim Application

The following example is about the treatment of a claim, CL, issued by a submitter, SU, to an Insurance Company, IC, concerning an incident, IN, involving an amount AM of money, on a trip, TR booked with Travel Agent: TA. The claim is treated in IC by an employee AP with role approver.

The structure of all types involved is (roughly) defined as follows:

type person **is_a** thing
has_a name
has_a address
has_a bank

type employee **is_a** person
has_a salary
has_a function

type claim **is_a** thing
has_a submitter
has_a trip
has_a incident
has_a amount

type submitter **is_a** person
type ICemployee **is_a** employee
type approver **is_a** ICemployee
type expert **is_a** ICemployee
type cashier **is_a** ICemployee

type TAemployee **is_a** employee
type travel-agent **is_a** TAemployee

First we describe the processes in natural language, using numbers identifying the actions for easy identification with the boxes used in the COLOR-X diagram of Fig.1.

1.1. SU sends a message containing trip TR and incident IN and the amount AM of money involved to the approver AP of IC, upon which:
2.1. AP makes the claim object CL and asks the travel agent TA to verify the claim (possibly) within one week.
3.1. TA tries to verify CL within one week and return the answer to AP; if she is not able to do so:
2.2. AP decides, when TA does not answer, or when the answer of TA is "not OK" that the claim has not been not approved and informs SU accordingly.
2.3. AP decides, upon receipt of TA's answer, that the claim is less than $100 and asks the cashier CA to transfer the money to SU and informs SU accordingly.
2.4. AP decides, upon receipt of TA's answer, that the claim is larger than $100 and asks the expert EX to verify the claim and informs SU accordingly.
4.1. EX treats CL and informs AP about the result.
2.5. When the claim is approved AP fills in the determined amount in the claim and asks the cashier to transfer the amount of the claim to SU's bank account; when it is not approved AP informs SU accordingly.
5.1. The cashier CA orders the SU's bank to transfer the amount to SU's account.

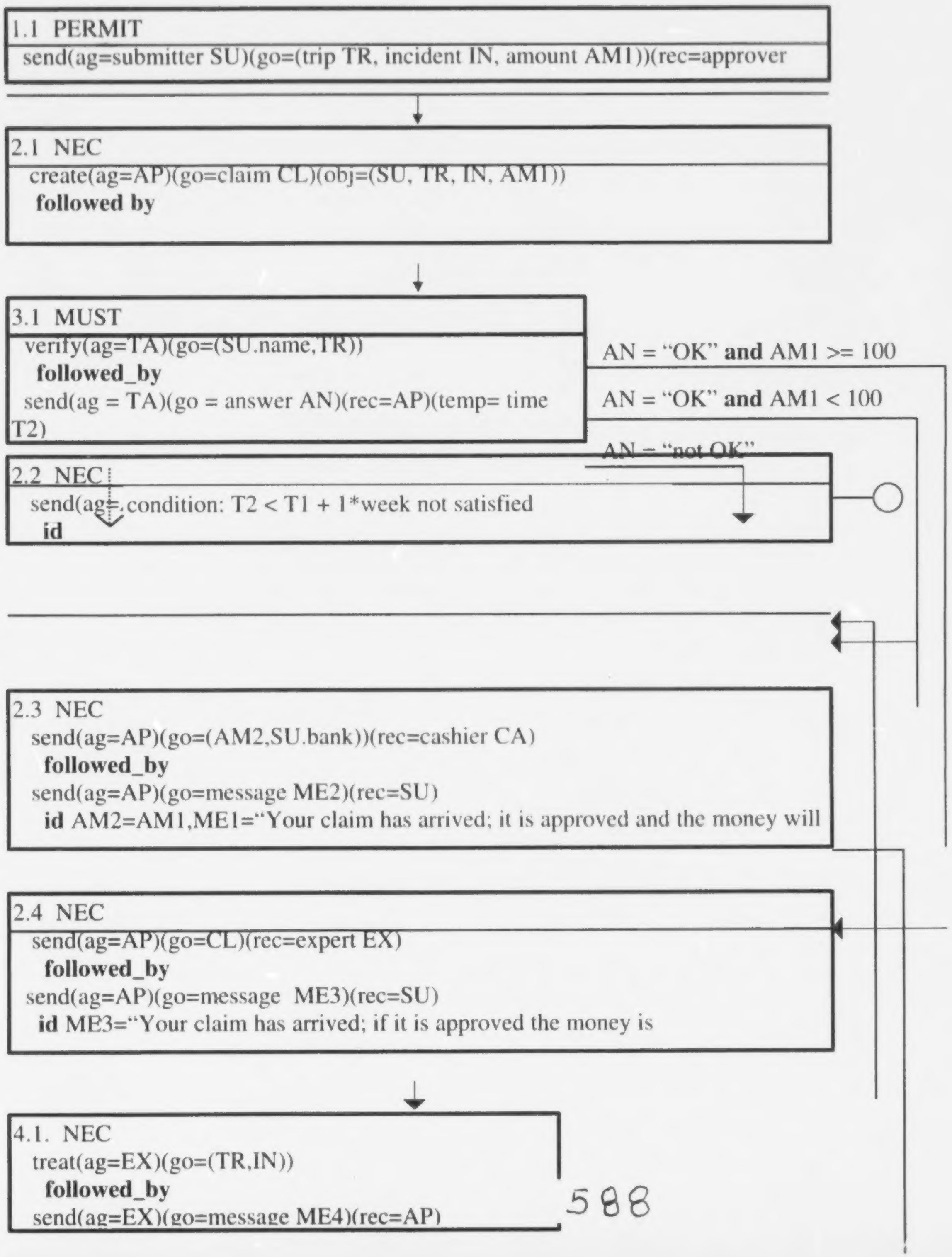
Note that a submitter SU occurs in three ways in our example: the first is active submitting the claim by sending a message to the approver of the bank, the second is as attribute in the claim; the third one is as receiver of a message. That such is possible is the benefit of considering alter egos as ordinary objects. In COLOR-X we have a diagram depicted in Figure 1, specifying the same processes. Note that this diagram is somewhat more detailed in the following respects:

* each box of actions has a mode: PERMIT, NEC or MUST. The latter one means an obligation based on some negotiating in the past: as we are not sure that the action is actually carried out within the prescribed time it is

necessary to define a counter measure. The mode NEC means: we can be sure the action is carried out. PERMIT is self evident.

* the actions are described in a formal language involving the participants and their roles;

* the objects involved are specified according to what parts of them are actually used; this is important for deriving the S&P rules.



1.1 PERMIT
send(ag=submitter SU)(go=(trip TR, incident IN, amount AM1))(rec=approver

2.1 NEC
create(ag=AP)(go=claim CL)(obj=(SU, TR, IN, AM1))
followed by

3.1 MUST
verify(ag=TA)(go=(SU.name,TR))
followed_by
send(ag = TA)(go = answer AN)(rec=AP)(temp= time T2)

2.2 NEC
send(ag=
condition: T2 < T1 + 1*week not satisfied
id

2.3 NEC
send(ag=AP)(go=(AM2,SU.bank))(rec=cashier CA)
followed_by
send(ag=AP)(go=message ME2)(rec=SU)
id AM2=AM1,ME1="Your claim has arrived; it is approved and the money will

2.4 NEC
send(ag=AP)(go=CL)(rec=expert EX)
followed_by
send(ag=AP)(go=message ME3)(rec=SU)
id ME3="Your claim has arrived; if it is approved the money is

4.1. NEC
treat(ag=EX)(go=(TR,IN))
followed_by
send(ag=EX)(go=message ME4)(rec=AP)

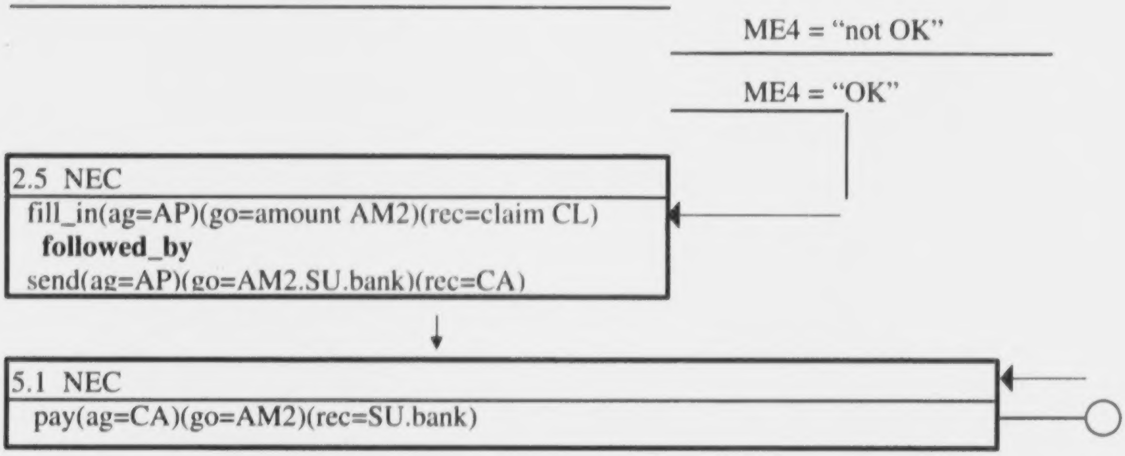

ME4 = "not OK"
ME4 = "OK"
2.5 NEC
fill_in(ag=AP)(go=amount AM2)(rec=claim CL)
followed_by
send(ag=AP)(go=AM2.SU.bank)(rec=CA)
5.1 NEC
pay(ag=CA)(go=AM2)(rec=SU.bank)

Figure 1. The Work Flow Diagram for the claim problem

4. Deriving Security and Privacy rules from WFM diagrams

We now derive the authorization tuples from the diagrams above. We use the following heuristic rules:

1. If an action involves data in a database, the agent of this action should be authorized to perform the corresponding actions on the database.

2. An action, with modality MUST, involves an obligation to perform a specific action within a prescribed amount of time. This implies that in some database, oblDB, the administration about this obligation is kept. The object which creates the MUST action, i.e. the agent of the action leading to this MUST action, can move the deadline (only shift it to the future of course); that is explicitly not allowed to the object who has to carry out the action. Of course this object can refuse to carry it out, but then penalties may be the result.

The databases involved are:

for IC: IC-claimDB, for TA: TA-tripDB, for obligations: oblDB

The syntax we will use:

AUTH <name database, role actor, operation, allowed or not allowed>

The numbers refer to the diagram.

2.1 AUTH<IC-claimDB, approver, create, allowed>
3.1 AUTH<oblDB, approver, shift, allowed>
AUTH<oblDB, travel_agent, shift, not allowed>
AUTH<TA-tripDB, travel_agent, read:submitter.name, allowed>
AUTH<TA-tripDB, travel_agent, read:trip, allowed>
4.1 AUTH<IC-claimDB, expert, read :trip, allowed>
AUTH<IC-claimDB, expert, read :incident, allowed>
2.5 AUTH<IC-claimDB, approver, write :amount, allowed>
5.1 AUTH<IC-claimDB, cashier, read:amount, allowed>

5. Roles, Types and Protection in Mokum

In section 3 we have identified roles with subtypes. Roles give certain rights, so it is important to protect the distribution of subtypes. However, in Mokum any object can give itself or any other object any (existing) type. Only the usage of attribute values can be protected, not the usage of types (for more information see[van de Riet&Beukering94]). Seemingly the choice of subtypes for roles is a bad one and it would be better to attach a role as a value to an attribute of an alter ego. This would be a pity as it is so natural to represent roles by means of types, in particular while types have property inheritance and behaviour.

As a matter fact, also in reality a person can put on different guises. For example, he can try to do as if he is a bank employee so that he can raise his own account. The way protection is provided in Mokum is by means of collections, which have to be protected by their collection keepers. In the case of the above person there is a collection of (bank) employees, CBE, kept by a special object, Bank Administrator, BA. A simple protection rule is that an object can become a member of CBE only if the request comes from a manager of the bank. So a message has to be sent to BA by another object in the role of manager. Again there is a protection problem here for who can control whether that object is really a manager and not a fake one? The answer is simple: BA not only controls who is employee it also has a collection of managers. So, the script for BA could contain the following code:

```
    trigger: become_employee
        /* check if sender is some manager and member of collection of managers */
        add person to CBE
```

In short, protection on roles is established as follows: any alter ego wishing to have a certain role can give itself the corresponding subtype. This gives the alter ego access to obtaining properties such as attributes and scripts. Only when some collection keeper accepts the alter ego, it is accredited. It is comparable to the way the title of Professor is protected in Holland: it is not protected and everyone can call himself a Professor; only by asking the University whether a certain person is Professor, one can be sure. Because protection is provided using a combination of syntactic checking (private attributes can appear only at certain places) and semantic checking (being in a collection or not) the keepers of the collections play key roles. When we deal with Cyberspace we have several more or less autonomous subsystems, such as in our example:

* the travel agent who has its own databases for access rights kept by a Travel Agent Administrator TAA,
* the Insurance Company with Administrator ICA,
* the Bank with administrator BA and
* the individual person who can submit a claim.

In the last case one can imagine that protection for individuals is provided by special protection agencies as it is also is the case for protection of properties such as a house.

6. Implementing the WFM example and the Authorization tuples in Mokum

We shall now discuss how an architecture can be designed, based on Mokum protection primitives, guaranteeing security and privacy. We assume that in Cyberspace local systems S exist each with an Administrator SA. All inter system protection problems are being solved by the cooperating SAs. Example: when the approver of IC sends a request to TA to check the claim submitted by the Submitter SU, he actually sends a message to ICA. This ICA first checks whether the message indeed comes from AP, and not by a fake approver, e.g. SU himself. Only when ICA is convinced that the approver is indeed an employee of IC and has the corresponding function (role) he sends a message to the TAA involving the request about the claim. Now it is TAA who has to check that the sender of the message is indeed ICA and not someone else, such as SU. Here we are dependent on the security properties offered by the Net. Also the message from SU about the claim can be treated in this way so that AP knows that the message indeed comes from the person identified by the sender of the message and not from someone else trying to make a joke. The question whether these SAs can themselves being trusted can be split up in trust in local protection and in global protection. Local protection is guaranteed in so far the code of the Administrators can be trusted. For the global protection we are dependent on the security provided by the Net. We refer to [Olivier95] for an architecture based on secure communica-ting of federated databases. Let us now see how the WFM rules can be represented in Mokum code in which protection is guaranteed. Let us take as example the Approver AP:

```
type approver is_a IC_employee
        has_a ....
    trigger claim_submission:
        /* peels out submitter SU, trip TR, incident IN and amount AM from the       */
        /* message and creates the claim CL; determines travel agent TA from TR      */
        /* makes message M from SU.name and TR.                                      */
        set_timer no_answer_from_travel_agent
```

Note that set_timer and drop_timer actually call for some Mokum Administrator who uses the database OblDB in which the obligations are stored and which sends off time_triggers at the appropriate times to the object who set the timer. In this case AP.

```
        send(TA,M)
    end_of_trigger claim_submisson
    trigger answer_from_travel_agent
        drop_timer no_answer_from_travel_agent
        CL.amount < 100 ....or
        /* make message M from CL                                                    */
        send(expert, M)
    end_of_trigger answer_of_travel_agent
    trigger no_answer_from_travel_agent
        /* make message M for SU with "Sorry..."                                  */
        send(SU, M),  ...
    end_of_trigger no_answer_from_travel_agent
```

We now turn our attention to the expert. In actual practice an expert has a collection of tasks which is filled by his manager and emptied by the expert himself at times determined by the amount of work and the availability of the expert. In another paper [van de Riet&Burg&Gudes&Olivier97] we treat this more complicated example. In our case we simply assume that the expert is always ready to treat a task. We also assume that it is not an autonomous system which can do the job, but it has to be the person himself whose expertise is needed.

```
    type expert is_a ICemployee
        proc treat(claim)
        /*actually here the real human expert is called to do the job              */
    trigger get_a_claim
        /*Determine claim CL from message                                          */
        treat(CL.trip, CL.incident),
        /*Make a message M from result of investigation:                           */
        send(sender, M)
    end_of_trigger get_a_claim
```

Let us now see how the authorization rules, so carefully derived in section 4, will be dealt with. For example the rule:

AUTH<IC:claimDB, expert, read:trip, allowed>

Evidently, this rule implies that a database is kept with all the claims maintained by the administrator ICA. If we assume that the details of the claim are stored in this database and that all references to a claim in the Mokum code are real references, or object identifiers to objects in this database, the above rule not only makes sense, it is also very appropriate. The above implementation shows that the expert gets pointers to CL.trip and CL.incident. If he now wants to see the values of these objects, he has to ask permission to do this explicitly to ICA. ICA checks whether this expert is real (accredited one) by looking in its collection of accredited experts **and by looking in its list of authorization tuples**. A good question is: if the expert, according to the above implementation, only gets the identifiers of CL.trip and CL.incident, why should this be checked also by the ICA. The answer is: in actual practice the expert might have got the identifier of CL inadvertently (as also suggested by the WFD of Fig.1. box 2.4).

Of course all this is been done automatically as soon as the expert wants to see the claim. All this is under the assumption that all claim objects are in the claimDB and that access is controlled by one administrator. For the travel agent to see his part of the claim the situation is somewhat more complicated because that part is at another site. The administrator at his site can see from the request that the value of the object is somewhere else. So there must be communication between this administrator TAA and ICA.

7. Conclusion

In this paper we have shown an architecture how to obtain S&P rules from a specification using modern Workflow management tools. Also it has been discussed how programs can be derived automatically which maintain in a guaranteed way these S&P rules. What we have not discussed is how these S&P rules can be maintained globally in a guaranteed manner. For space reasons this is being done in a parallel paper [van de Riet&Burg&Gudes&Olivier97].

References

[Atluri&Huang96] V.Atluri, W-K.Huang, "An Extended Petri Net Model for Supporting Workflows in a Multilevel Secure Environment", "Proc. 10th Annual IFIP WG 11.3 Conf. on Database Security, Como, Italy, August, 1996, pp. 199-216..

[Burg96] J.F.M.Burg, Linguistic Instruments in Requirements Engineering, IOS Press, Amsterdam, 1996.

[Georgakopoulos95] Georgakopoulos, D., Hornick, M., andSheth, A., ``An overview of workflow management: from process modeling to workflow automation infrastructure," Distributed and Parallel Databases, Vol 3, No. 2, 1995, pp. 119-154.

[Olivier95] Olivier, M.S., ``Self-protectingObjects in a Secure Federated Database", in DL Spooner, SA Demurjian and JE Dobson (eds), Database Security IX: Status and Prospects, 27--42, Chapman & Hall, 1996.

[Olivier96] Olivier,M., S. ``Using workflow to enhancesecurity in federated databases," Proc. 2nd Int. Conf. on Communication and Multimedia Security, Essen, Germany, 1996, pp. 61-72.

[van de Riet&Beukering94] R. P. van de Riet and J. Beukering."The integration of security and integrity constraints in Mokum". In J. Biskup, M. Morgenstern, and Landwehr, C. E., editors, Database Security, VIII, Status and Prospects, pp. 223-246, IFIP, North-Holland, 1994.

[van de Riet,Burg96a] R. P. van de Riet and J. F., M. Burg,``Modelling alter-egos in Cyberspace: who is responsible?", Proc. WebNet96, AAA, 1996.

[van de Riet&Burg96b]Riet, R.P. van de, J.F.M.Burg, Linguistic Tools for Modelling Alter Egos in Cyberspace: Who is Responsible? Journal of Universal Computer Science, Vol 2, number 10, Springer1996.

[van de Riet&Gudes96] R. P. van de Riet, and E. Gudes,``An object-oriented database architecture for providing high-level security in Cyberspace,"Proc. 10th Annual IFIP WG 11.3 Conf. on Database Security, Como, Italy, August, 1996, pp. 92-115.

[van de Riet&Junk&Gudes97] R. P. van de Riet, A. Junk, and E. Gudes, ``An object-oriented database architecture for providing high-level security in Cyberspace," Data and Knowledge Engineering, Elsevier, 1997.

[van de Riet&Burg&Gudes&Olivier97] R. P. van de Riet, J.F.M.Burg, E. Gudes, and M.Olivier, Modelling Security and Privacy in Cyberspace using Work Flow Management, submitted for presentation, 1997.

Organising Distance Learning Process thanks to Asynchronous Structured Conversations

Claude Viéville
Frédéric Hoogstoël
email: {Claude.Vieville, Frederic.Hoogstoel}@univ-lille1.fr
Laboratoire TRIGONE
Université des Sciences et Technologies de Lille
Cité scientifique
59655 Villeneuve d'Ascq CEDEX, France

Abstract: This paper describes a communication system accessible by a Web browser. The main advantage of this system is to encourage a collaborative way of learning using asynchronous communication channels. The conversation is strongly structured by the system itself which helps the users to co-ordinate their actions playing their respective roles within a task. A conversation always occurs in the context of a task where each user plays a particular role. The system is built around the notion of active form which is the single way for the user to communicate with the system. This system can be used by the actors of the educative process to organise their work.

1 Introduction

This paper describes an asynchronous collaborative learning system which aims to support a distance education process on the Web. What attracts the Internet to an educational institute is a large communication network to exchange information in two ways, the on-line browser and the courseware package distribution. So the challenge we have to face is to change information exchanges into learning activities. For this reason, we are interested in second generation servers which respond better to educational needs : better interactivity between video-clip, text, images, and so on ; enabling re-use of all the supports we have developed in a fully integrated manner ; inclusion of graphics and formulae is compulsory for a lot of curricula ; embedded courseware corresponds with the multiplicity of training pathways for individualised training and the ease of navigation required. As a minimum requirement, the system needs communication facilities to enhance real collaboration between users and tutors. In the EONT project[1], in which we are participating, we are verifying these hypotheses. And in the DEMOS project[2] we are designing, developing the asynchronous communication system presented more precisely in this paper. To develop our system we distinguish three spaces in which the activities of learners take place : information space, action space and communication space. The communication space depends on the institute, and organises the interactivity between the different spaces to correspond to a pedagogical practice. After a short introduction of the application field, the paper presents the functional specification of the system we are currently testing.

2 Educational Context

The CUEEP (Centre Universite-Economie d'Education Permanente) is an institute of the University of Sciences and Technologies of Lille in northern France which is concerned with several activities: further education for adults, research into educational engineering (open learning and new communication technologies), transfer within the context of new technologies in education. Some experiments of the co-operative system Co-learn had been set up during these two last years. Now we search to integrate this communication system into our distance education organisation. To continue our work of research into the use of tools of communication in distance education we are conducting a project to deliver course on the Web based on collaborative learning. This project is mainly supported by the European Commissions through the Telematics for Education programme. In this framework, we are setting an Asynchronous Collaborative Learning System in the DEMOS project. This system relies on a second generation of Web server (HyperWave from University of Graz - Austria).

---

[1] An experiment in Open Distance Learning using New Technologies - part of the Socrates programme of the European Commission

[2] Distance Education and tutoring in heterogeneous teleMatics envirOnmentS - part of the Education and training programme of the European Commission

3 Overview of the Services

This asynchronous communication system will provide a set of services from the same family as those already provided by electronic mail (email), electronic forums (forum), Bulletin Board Systems (BBS) and the News. Its ambition is to give users real help with their *tasks* by avoiding several well-known drawbacks of current systems [Terry 1991] and to propose a structuring of the conversation so that it is very efficient to communicate and collaborate via such a system [Vieville 1995]. The measurement of the efficiency of this system could be made upon the following :

- time-saved during the co-ordination phase of a collaborative process [Bussler and Joblonski 1994],
- time-saved when reading each others contributions,
- enhancement of the quality of arguments produced during a debate[Desaranno and Put 1994],
- better involvement of users in the collaborative processes.

The Co-Learn project is an important input to the specification of such a system. In [Derycke & al 1992], the interest of developing Collaborative Learning activities has been explained. It is outside the scope of this document to argue in favour of educational processes which are based on collaboration between learners and tutors. In [Kaye 1995] it is also written as a result of the Co-Learn project, that "*it might have been preferable to put emphasis on the Asynchronous Communication mode as the basic substrate for communication between learners and tutors... In this way the Asynchronous Communication Mode would provide the glue which would hold a course together, inter-linking the real-time sessions, and providing a forum for continuing discussion and collaboration after each of these sessions.*" The reader who is interested by this discussion will find pertinent papers on this subject in the reference section [Harasim 1993], [Henri and Rigault 1996], [Kirsche & al 1994]. Jonassen, in [Jonassen 1996], gives an excellent overview of the possibilities of Computer Mediated Communication (CMC) in educational process.

3.1 Basic Services

The ACLS offers a set of basic services enhanced by a subset of complementary services which are needed to manage, adapt and integrate the system using existing communication tools to meet users' needs [Palme 1992], [Palme 1993], [Turoff 1991]. Globally the basic services provided by this asynchronous communication system are:

- informal exchanges,
- question-answer exchanges,
- date negotiation [Woitass 1990],
- pro-con argument production,
- action negotiation [Rogers 1995]
- opinion collection.

Each of these services could involve people regardless of the context of a collaborative task, or be used in the framework of a task process involving the group. In this latter case the exchange is automatically classed as public, unless specifically defined as private. The task in which the communicators are involved in is very fundamental as it will define the context in which the exchange has occurred [Ellis and Wainer1994].In this

ACLS, electronic mail is not distinguished from electronic forums or news systems as a means of communicating between people. The ACLS provides an integrated view of exchanges whatever channel is used (i.e. email, forums, news, BBS etc.) [Benford & al 1992].

This basic service will allow the members to select, fill in, edit, and submit a form which will complete an exchange. Exchanges are linked to each other by a temporal relation. The creation of a new exchange is a particular case of the creation of a contribution which becomes the root of the exchange. The ACLS also proposes other complementary services to its basic services. These will be described in the following section.

3.2 Complementary Services

To encourage co-operation ACLS will provide a service which gives information on its *users*. The communication needed by users during the task process will be supported inside a *group activity*. The group activity is the context in which the *exchange* of a communication occur. One and only one organisational group is attached to a group activity. The exchanges of a communication are structured sets of contributions. Each exchange is regulated by a set of global rules pre-defined at the installation of the ACLS. This set of rules depends on the way people of the organisation work together [Vieville 1995]. Obviously default rules are proposed during the installation phase. To participate in a group activity a user needs to be added ; he then becomes a *member* of the group activity.

It is also possible task by task to create subgroups in which all the members play an identical role with regards to the aim of the task. For example, if a collaborative writing task is started, subgroups of "authors", "editors", "reviewers" are created by the initiator of the task. Belonging to a subgroup will give different rights to the objects in the ACLS. A search service is available for all the users who want to find any objects in the ACLS. Users, group activities, sub-groups, forms, exchanges and tasks are searched and displayed to the user of the search service. To start a search operation, the user must fill in fields of a search form. The user has to define in the form which criteria the search should use. It is possible to search on the attributes and/or the contents of any types of objects of the ACLS. Authorised users will use the administration service to create/modify attributes ; delete/archive/open/close user and group activities. This administration is done by filling in an administrative form. Users are added and removed from group activities by using the registration service. A subset of authorised users with appropriate rights will have access to this service. Registration is performed by filling out a registration form. Only when a group activity has appropriate parameters may a user register himself for that activity.

A service of notification allows members, who have subscribed, to be notified when something is appended to the group activity. Filling in a notification form is the proposed way to subscribe to the notification service. The notification service allows to the user to receive (or avoid reception of) the events generated inside the ACLS. The kinds of events are :

- "group activity" list has changed,
- list of users of the ACLS has changed,
- status of a group activity has changed,
- list of tasks for a particular group activity has changed,
- list of exchanges for particular tasks has changed,
- list of forms for a particular exchange has changed,
- a deadline relative to a task is going to arrive,
- a deadline relative to a task has been detected,
- a particular user activity has been detected,
- a particular group or subgroup activity has been detected.

The events are sent to the notification recipient which could be an electronic mail address, a news group, or another task of any other group activity.

4 Several Implementation Key Points

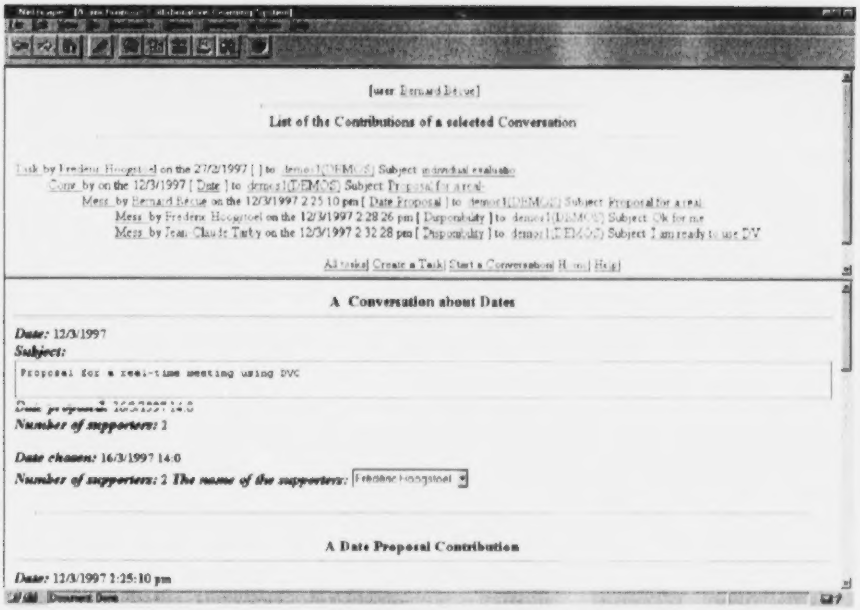
From the implementation point of view, ACLS relies on the architecture of an open system of CSCW called ODESCA[Hoogstoël 1995] which is built on the integration of an activity server using an object database for persistency with a WWW information server.

The access to the ACLS functions is realised by the way of the CGI mechanism of a web server. The CGI interface takes in charge the management of the transactions which is not supported by the web servers. This interface is also in charge of the management of the templates database of forms according the organisation and the users. Finally, it also communicates with ODESCA to obtain the conversation state, the list of types of templates allowed for a contribution and other functions less specific to asynchronous communication activities as the information on group members. This CGI application continuously updates a database where the interactions between users and ACLS are stored in order to give information to measure usability of the system. The data forwarding from the user station and CGI application is done according the HTTP protocol. This protocol does not support transaction by itself, so a mechanism has been designed to reject non valid request which has already been submitted. For example, we must avoid a user to submit the same form several times when he uses the moving back functionality of a web browser. A user thanks to a standard web browser of the Internet is able to get the list of the tasks in which he is involved in. Then, using the navigation functionality, he can get the list of the conversations of a selected task. Finally, he will get the list of the contribution of a particular conversation. A synthetic view of the state of the conversation remains always accessible as well as the set of the contents of all the contributions of a conversation [Fig 1].

Each time a user wishes to add a contribution at the heart of ACLS, the ODESCA server activates itself to propose him the list if the types of forms which are accessible. This list is computed by taking into account the state of the conversation in which the user wishes to converse, according to the role of the user and according to the kinds of the contributions he has already submitted. For example, in a conversation to define a date, the initiator of this conversation will receive from the ODESCA server a list of two forms : using the first one he will be able to convoke the persons at a date selected by the members of the group ; with the second one he will be able to announce the abort of the meeting for any reason. The submission of one or the other form will finish the current conversation. In this same conversation, all the other members of this group will receive from ODESCA a form in which he will indicate if the date is convenient for him.

ACLS makes a clear distinction between the presentation objects seen and manipulated by the user and the objects manipulated by itself. When a user creates a new object (i.e. new task, new conversation...) the system selects appropriate list of templates and the user has then to select one of these. Then, he has to fill the fields of this template. The templates is a HTML form controlled by javascripts. Javascript controls user input date for each field whose content is interpreted by the system. As the templates are semi-structured messages, some fields are not interpreted by the system but just stored and some other ones needs a strict control. Before being submitted a form which carry all the data of the template is locally controlled by a javascript. Designers of these templates encounters difficulties due to the lack of standardisation of javascript among browser. Netscape currently presents the most advance feature as it is able to manipulate HTML objects such as select object.

The current implementation takes in charge several parameters suitable for the organisations in which ACLS is used but also several other one suitable for the users. An organisation can select among an existing template database of forms but also edit its own database. The ACLS system uses HTML documents and proposes an extension which allows itself data on the flow according to the sate of the conversation or the role of the user. The edition can be done by anybody knowing a HTML editor and the meaning of the variables of the ACLS system. By using a modification process, it is very easy to realise a new templates database in another language. This option is also proposed for user by user. It can be used to reduce the complexity of a given set of information according to the skill of the users with the system. As the models are stored in the HTML format, a classic web browser such as Netscape Navigator or Microsoft Internet explorer can used to access to the ACLS system. This choice allows a large usage of the ACLS.


[user Bernard Bécue]

List of the Contributions of a selected Conversation

Task by Frederic Hoogstoel on the 27/2/1997 [ ] to demos1(DEMOS) Subject individual evaluatio
Conv by on the 12/3/1997 [Date] to demos1(DEMOS) Subject Proposal for a real-
Mess by Bernard Bécue on the 12/3/1997 2 25 10 pm [Date Proposal] to demos1(DEMOS) Subject Proposal for a real
Mess by Frederic Hoogstoel on the 12/3/1997 2 28 26 pm [Disponibility] to demos1(DEMOS) Subject Ok for me
Mess by Jean-Claude Tarby on the 12/3/1997 2 32 28 pm [Disponibility] to demos1(DEMOS) Subject I am ready to use DV

All tasks| Create a Task| Start a Conversation| Home| Help|

A Conversation about Dates

Date: 12/3/1997
Subject:
`Proposal for a real-time meeting using DVC`
Date proposed: 16/3/1997 14:0
Number of supporters: 2

Date chosen: 16/3/1997 14:0
Number of supporters: 2 ***The name of the supporters:*** Frederic Hoogstoel

A Date Proposal Contribution

Date: 12/3/1997 2:25:10 pm

Figure 1: View of a Conversation to agree on a date

On the other hand, it remains possible to integrate the ACLS functionality at the heart of another application. As a matter of fact, It exist a templates set in a MIME format which offer a mean to implement a new interface less general than a web browser and so more adapted to a specific context of work. The implementation can be done in any language as ACLS interface is just a definition of a protocol. JAVA seems to be a good candidate to this implementation.

The notification mechanism which is in currently in test but soon available allows users to never consult the ACLS. They only have to let a email agent active on their station. This agent will receive a notification message coming from ACLS telling them what is new in ACLS for them. A backward link helps them to directly consult the task and the conversation which includes the major events.

5 Conclusion

A particular attention has been paid in the methodology of design in order to work with the user group. This system has been designed incrementally ; it means that, rapidly, with only a few functions it has been usable by the members of the user group who sent feedback to the designers. This participative approach has certainly given to this system a good level of usability. At the moment this paper is written, implementation of the first release of the prototype is finished and results of usability are soon available.

6 References

[Benford & al 1992] Benford S., Turroff M. and Palme J. (1992), An ISO standard to support asynchronous group communication. In: Computer Standards & Interfaces Elsevier Science Publishers B.V. (North-Holland); 14:363-373

[Bussler and Joblonski 1994] Bussler C. and Joblonski S. (1994), Implementing Agent Coordination for Workflow Management Systems Using Active Database Systems. IEEE software; 4:53-59

BEST COPY AVAILABLE

[Derycke & al 1992] Derycke A. and al. (1992), Representation Models for Collaborative Educational Situations and Collaborative Learning activities, Deliverable 4 of Co-Learn D2005

[Desaranno and Put 1994] Desaranno S. and Put F. (1994), Co-ordinating group communication. In: Computer networks and ISDN Systems, Elsevier Science B.V., pp S129-S138

[Ellis and Wainer1994] Ellis C. and Wainer J. (1994), Goal-based models of collaboration. In: Collaborative Computing, Chapman & Hall, pp 61-86

[Harasim 1993] Harasim L. (1993), Collaborating in Cyberspace : Using Computer Conferences as a Group Environment. Interactive Learning Environment; 3(2): 119-130

[Henri and Rigault 1994] Henri F. and Rigault C. (1994), Collaborative Distance Learning and Computer Conferencing In: Advanced Educational Technology: Research Issues and Future Potential (T.Liao ed.) NATO, ASI series, Springer Verlag , pp 45-76

[Hoogstoël 1995] Hoogstoël F. (1995), Une approche organisationnelle du travail coopératif assisté par ordinateur. Application au projet Co-Learn, thése de Doctorat en Informatique à l'université des Sciences et Technologies de Lille

[Jonassen 1996] Jonassen D. (1996), Computers in the classroom ; mindtools for critical thinking. Prentice Hall

[Kaye 1995] Kaye A. (1995), Final Evaluation Report, Synthesis of Co-Learn Trials and Experimentation, Deliverable 33 of Co-Learn d2005

[Kirsche & al 1994] Kirsche T., Lenz R., Ruf T. Wedekind H. (1994), Co-operative Problem Solving using Database Conversations. IEEE software; pp 134-143

[Malone 1987] Malone T. (1987), Semi-Structured messages are surprisingly useful for Computer-Supported Collaboration. ACM trans. On Office Information Systems; 5:115-131

[Palme 1992] Palme J. (1992), Computer Conferencing Functions and Standards. In: Collaborative Learning through computer conferencing, Kaye A.R., pp 225-245

[Palme 1993] Palme J. (1993), Standards for asynchronous group communication. Computer communication; 16-9:532-538

[Rogers 1995] Rogers (1995), The Use of an automatic "TO Do " list to Guide Structured Interaction.. In: Proceedings of CHI 95, Denver, Colorado, USA, pp 232-233

[Terry 1991] Terry D. (1991), 7 steps to a better mail system. In: Message handling Systems and Application Layer Communication Protocols, Elsevier Science Publishers B.V. (North-Holland), pp 23-33

[Turoff 1991] Turoff M. (1991), Computer-Mediated Communication Requirements for Group Support. Journal of Organizational Computing, 1:85-113

[Vieville 1995] Vieville C. (1995), Structuring conversation in asynchronous communication systems to support collaborative learning at a distance. In: Proceedings of ED-MEDIA 95, Graz, Austria, pp 816-817

[Woitass 1990] Woitass M. (1990), Coordination of Intelligent Office Agents Applied to Meeting Scheduling. In: Multi-User Interfaces and Applications, Elsevier Science Publishers B.V. (North-Holland), pp 371-387

Incorporating Shared Relevance Feedback into a Web Search Engine

Kevin Y. Wei, Microstrategy, Inc., USA, wei@strategy.com

Gregory C. Sharp, Computer Science Department, University of Wisconsin-Madison, USA, greg@cs.wisc.edu

Abstract: In this paper we introduce the *Way Sharp* system, a web search engine that uses shared relevance feedback together with corpus statistics to score documents and resolve queries. Passive feedback is gathered as the number of times a link is visited for a given query. Users may also actively update the database with relevance feedback. Queries to the database return a list of links that are scored using the relevance feedback information together with the information from the corpus statistics.

1 Introduction

The World Wide Web is a vast information repository, with 1.3 million servers and 95 million users [Wizards 96] [ISC 96]. Finding the right documents is not always easy. Consider a typical web search scenario, with our hero using the WWW as a virtual library for retrieving information about slime modes. First, she connects with a general purpose search engine and types some keywords in. The search engine parses her request, searches its database, and returns links, descriptions and scores based on information from within the document such as keyword statistics, META fields, and so on[AltaVista 96]. If the links returned by the search engine are of sufficient quality, our hero might start from one of them and wander around. In this process, she will visit many useful links which are relevant to her query. If the links are of uneven quality, our hero might read twenty or thirty links until she finds a informative link, or else she gets tired, hungry or bored. If the general purpose search engine fails, another search using the same query and same search engine will also fail. Modulating the query terms may work, but all too often this technique returns the same set of pages or else it returns a set of different unrelated pages.

Later, when a second person makes the same query with a similar goal, he or she will likely follow this same process of wandering around amongst the same set of links, modulating the same keywords and so on. If the precision of the search was low, knowledge about which links a previous user found useful may be useful for the second person. It may even be useful to know that previous users have made the same search, and found no links useful.

In our project, we developed a new, smarter search engine called *Way Sharp* . It is a search engine, so the initial stage is the same as other search engine in that it returns a sorted list of potential useful links. As users navigate through the returned links, the server keeps a record of all followed links as a passive measure of relevance feedback. Push buttons are provided by the system to let users explicitly add relevance feedback information. In this way, *Way Sharp* will become sharper and sharper as it learns from more and more users.

In a single user system, feedback about specific links can be used to bookmark the good links and to blot out the bad links (blotting out bad links is definitely a salve for battered nerves – GCS). In a multi user system, however, things are not so simple. For example, a single user might wish to provide a lot of negative information about a particularly hated link so that it never again appears as the result of a search. A society that judges free speech to be so highly valued, is difficult to justify a system that allows a single user to push any link off the list. One exception to this fairness rule might be within

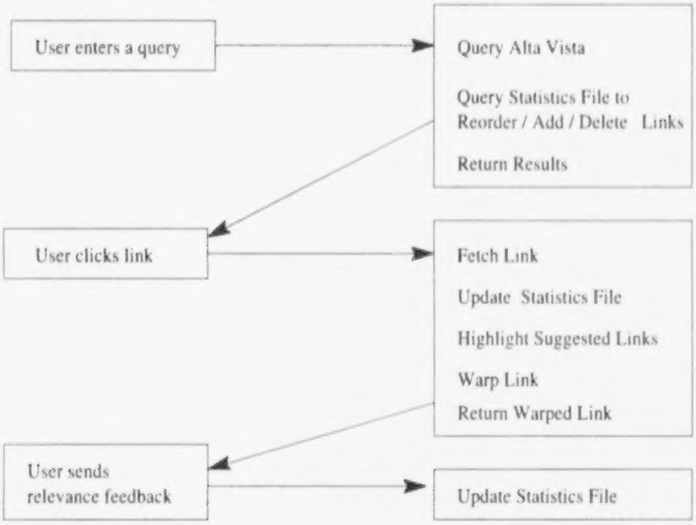


Figure 1: A Sample Query Session in *Way Sharp*

a trusted domain, such as within a single research group, where all opinions are of equal importance, and all opinions are trusted to be accurate. Outside of these trusted domains, such as in the internet, a retrieval system must restrict the amount of feedback that can come from any single user. Furthermore, the system must ensure that the feedback information is not allowed to dominate over other sources of information, such as corpus based statistics. Without some form of control, it is difficult to compare links which have relevance information with new links that have just been posted.

Another important aspect of information on the web is that it is highly changeable. As such, the value of feedback may be relative to when it was given. This is visible in highly evolving domains such as research or social groups. Answers to such questions as "What is the best way to code this algorithm?" or "Where is the best place to eat pizza?" change from week to week. This suggests that the weight given to relevance feedback might be better modeled as a function of time, which gradually decays. Under this model, relevance which is of a timeless nature must periodically reevaluated and reinforced.

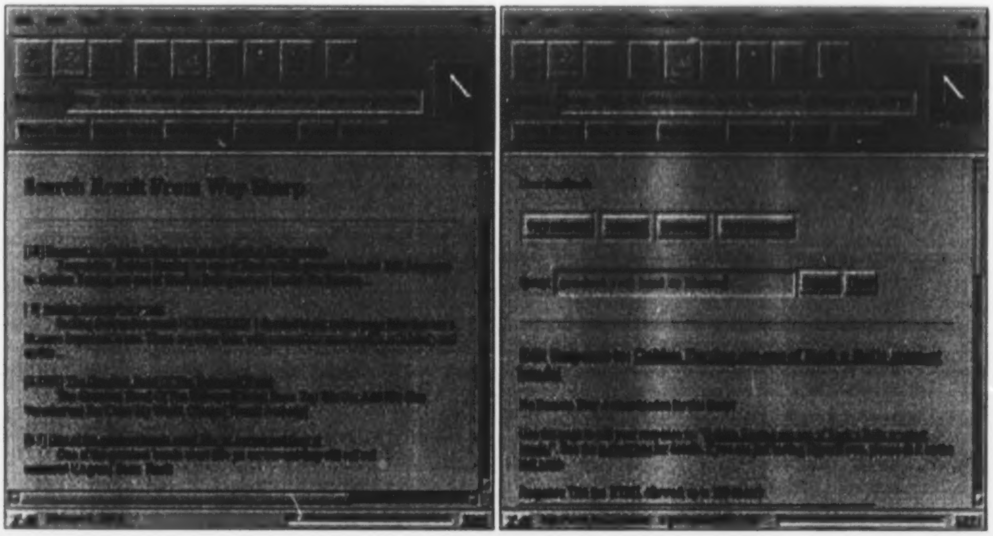
2 Way Sharp

The structure of our system is shown in [Fig. 1] which describes the sequence of actions that take place within a single session. Users begin a session by sending a query to the search engine, using a HTML form. By making a query, users "enter the system" in the sense that all future requests for web pages will pass through the *Way Sharp* server.

The server first calls Alta-Vista to get a number of matching links, then queries our local database to do three things. It reorders the links according to user feedback, adds some links as learned from users, and deletes some links if these links get negative feedbacks from users.

Clicking on a link sends a request to the system, which fetches the page, and records that the link was visited. Before the page is returned to the user, the links within the page are warped to point to the *Way Sharp* server. Some additional form controls are also attached to the top of the page so that users can provide feedback as they go. The server will accept requests to update the database from these relevance buttons. It also highlights the suggested links by blinking these links. A sample user page is shown in [Fig. 2].

At any time, users may send relevance information about the current page by pressing a button on the top of the web page. There is also a text entry field provided for the user to adjust the query. The user can exit at any time by opening a URL that outside of the system.



Search Result From Way Sharp

Figure 2: *Way Sharp* Search Results and Wandering the Web with *Way Sharp*

3 Implementation

3.1 CGI and Perl

The system use CGI for the front end interface and use perl as programming language. In addition to standard Perl function, we also used Libwww-perl and WWW-Search to implement Web related functions.

Libwww-perl is a collection of Perl modules which provides a simple and consistent programming interface (API) to the World-Wide Web[Hukins et al. 96]. The main focus of the library is to provide classes and functions that allow you to write WWW clients, thus libwww-perl said to be a WWW client library. It also contains some tools for parsing HTML, and tools of more general use.

WWW::Search is a collection of Perl modules which provide an API to WWW search engines[USC 97]. It currently supports AltaVista (web or news), Lycos, Yahoo and Hotbot. Currently WWW::Search includes the generic library, a back-end for AltaVista, AutoSearch (an program to automate tracking of search results over time), and a small demonstration program to drive the library.

3.2 Link Morphing

When user follows a link, it first goes to our server. The server extracts the name of the link and the original query from the input query string, and then updates the number of times that the link is visited. Next, the server fetches the page pointed by that link. Before give this page to user, server also have to morph all the links in that page to central server, highlight possible interest links and put relevance buttons on top of the page. Finally the server shows the page to user and the user can tell the server if the link is relevant by pressing the corresponding button.

Warped links need to carry along additional information and possibly the rank of the page. For example, if the original query is `pork+beans`, the link is `http://www.cs.wisc.edu` and the name of the *Way Sharp* server is:

`http://cgi.cs.wisc.edu/scripts/local.pl`

then the morphed link might be:

`http://cgi.cs.wisc.edu/scripts/local.pl?http=http://www.cs.wisc.edu&query=pork+beans`

BEST COPY AVAILABLE

Link	Query Keyword	Score
http://foobar.com/	foobar	1.36673
http://foobar.com/	piano	-1.22348
http://fubar.org/	foobar	-0.23424

Table 1: Main Database

Link	Query Vector	Score
http://foobar.com/	foobar+piano	0.16673
http://foobar.com/	foobar+fubar+foo	2.13173
http://fubar.org/	foobar+piano	-0.99484

Table 2: Query Vector Database

Relevancy information is generally not included in the warped link, as this information is sent by clicking a button within the FORM interface rather than by clicking on a URL.

3.3 Indexing

The perl front end communicates with the database subsystem, a back end application written in C++. The database subsystem maintains two databases, a main database and a query vector database. The main database is a single table that relates links with keywords and scores. For each link/keyword pair, only one score is kept. An example of this internal table is shown in [Tab. 1]. There is a problem with this format in that the scores are only maintained on a keyword by keyword basis. Information about the relationship between keywords of a given query vector is lost. In order to preserve this information, *Way Sharp* maintains a second database which relates link/query-vector pairs with relevance scores. An example of the query vector database format is shown in [Tab. 2].

3.4 Scoring

When users perform a search, the candidate links are scored and ranked. In order to produce an unambiguous ranking, the relevance feedback information must be combined with the statistical corpus information into a single numeric score.

The scoring system that *Way Sharp* uses is an ad hoc function that calculates the rank values based on previous relevance information. For each link/query vector pair, the following scoring function is used with an α of 10, and a β value of 1.

$$\frac{\alpha}{\text{rank}} + \beta\,\text{relevance}(\text{link}, \text{query}) \tag{1}$$

where rank is the ranking of the link in the returned link set when the underline search engine(it's Alta Vista in our case) is called.

The relevance function is the normalized sum of all relevance associated with the page for the keywords within the query.

$$\text{relevance}(\text{link}, \text{query}) = \frac{1}{n}\sum_{i=1}^{n} \text{rel}(\text{link}, \text{kw}_i) \tag{2}$$

Where n is the number of key words in the query.

The rel function is the relevance of a single keyword to a given link. This measure is simply the score stored within the database. Every time the database is updated, the scores are adjusted depending on the type of feedback, adding 2 for "very+relevant", 1 for "relevant", 0.1 for "no+opinion", -1 for "irrelevant" and -2 for "very+irrelevant". A link that was visited but not given explicit feedback is interpreted to having passive feedback and is added 0.1 to the score. So the value of relevance function

Figure 3: Enhanced Search Results Page

in equation(1) is changing as more and more users give relevance feedback and the actual ranking of each link will be different from the ranking returned from underline search engine.

As users enter more and more relevance information into the database, the scores within the database grow without bound, either positively or negatively. As this happens, links that have been marked relevant multiple times will dominate over links that are scored only based on a search engine ranking. This is a problem because newly created links cannot compete with the older links. In order to alleviate this problem, the relevance scores must be normalized over time. One method for doing this, borrowed from reinforcement learning techniques[Nilsson 96], is to degrade each score over time. Relevance information can be considered to be the rewards. Furthermore, we could also consider the corpus statistics as another kind of reward that is updated every time that a query is made to the search engine. The update function becomes:

$$\mathrm{rel}(\mathrm{link}, \mathrm{keyword})_t = \gamma \mathrm{rel}(\mathrm{link}, \mathrm{keyword})_{t-1} + R \tag{3}$$

Using the measures described above, R can be either a score between +2 and -2 (for relevance feedback), or else R could be a score between +.5 and +10 (for corpus statistics).

3.5 Networking Issues

When users makes a query using the search command, an HTTP request is sent to *Way Sharp* , *Way Sharp* sends a request to Alta Vista, Alta Vista returns a list of links, the list is adjusted and returned to the user. Similarly, when users wander through pages they are sending an HTTP request to *Way Sharp* , *Way Sharp* requests the page from its server, the server returns the page, the page is adjusted and returned to the user. Users must wait for at least two complete round trips, where they would only have had to wait one round trip if they search or wander without the aid of the system. Furthermore, there appears to be no recourse when using CGI, because the CGI script must generate and return a static web page.

In order to reduce the cost of supplying relevance feedback, we have developed an alternate user interface that allows a user to update the relevance for multiple pages in a single step. This interface, shown in [Fig. 3], lets users update all of the links for a given search page with a single network access. One drawback to the design is that it requires users to make an accurate judgement of relevance based solely on the abstract descriptions that are returned by the search engine.

BEST COPY AVAILABLE

4 Summary and Future Work

We have presented an implementation of a web search engine that takes advantage of user supplied relevance feedback in order to improve precision. Our system provides a simple HTTP interface to a shared database, with capabilities for making queries and updates either interactively through a web browser, or else in batch through a series of HTTP requests. Our system has small disk space and networking requirements because relies on general purpose search engines for corpus based statistics, and only stores the relevance information.

A problem with the current CGI implementation of *Way Sharp* is that it requires a minimum of two round trips for each page fetched. To improve network response time, CGI server update requests and link morphing may be removed from the server, and performed by the web client instead. A client code application written in Java would be able to send the CGI server updates in parallel with the page fetch. Additional network savings could be made by batching the update requests.

One way in which the quality of the returned links might be improved is by allowing users to see the relative feedback score separate from the corpus statistics score. This would allow users to decide for themselves whether or not they trust the opinions of previous users, and make decisions accordingly. To preserve a balance between corpus statistics and feedback, a ranking system could be developed that merges lists of pages that score high on either scores together with a list of pages that scores high on a combined score.

References

[Gundavaram 96] Shishir Gundavaram.(1996). *CGI Programming on the World Wide Web.* O'Reilly & Associates Inc.

[Joachime et al. 96] Thorsten Joachims, Dayne Freitag & Tom Mitchell.(1996). WebWatcher: A Tour Guide for the World Wide Web, Technical Report, CS Department, CMU, Pittsburgh, PA 15213.

[Nilsson 96] Nils J. Nilsson.(1996). *Introduction to Machine Learning,* Draft of a Proposed New Textbook, pp. 159-174.

[Wizards 96] Network Wizards, Inc.(1996). Internet Domain Survey, http://www.nw.com/zone/WWW/report.html

[GTRC 96] Georgia Tech Research Corporation.(1996). GVU's WWW User Survey, http://www.cc.gatech.edu/gvu/user_surveys/survey-10-1996/.

[ISC 96] Internet Solutions Corporation.(1996). Internet Statistics - Estimated. http://www.netree.com/netbin/internetstats

[AltaVista 96] AltaVista Technology, Inc.(1996). AltaVista, http://www.altavista.com/

[Hukins et al. 96] Tom Hukins, Gisle Aas & Otis Gospodnetic.(1996). Guide to libwww-perl for Perl5, http://www.york.ac.uk/ twh101/libwww/

[USC 97] University of Southern California - Information Sciences Institute.(1997). WWW::Search, http://www.isi.edu/lsam/tools/WWW_SEARCH/index.html.

Acknowledgements

The authors would like to thank Prof. Belew for his productive discussions with us and the first author would like to thank Benjamin Li and Charles Veley in MicroStrategy Inc. for their support.

Cultural Implications in the World Wide Web: A Case for Research

Martyn Wild
Faculty of Business, Edith Cowan University
Perth, Western Australia 6018
m.wild@cowan.edu.au

Lyn Henderson
School of Education, James Cook University
Townsville, Australia 4811
lynette.henderson@jcu.edu.au

Janice Burn, Ron Oliver, Sue Stoney
Edith Cowan University, Perth, Western Australia 6018

Abstract: The World Wide Web is being used increasingly by universities to carry interactive means of teaching and learning, often as part of a more general policy towards flexible or open learning. This paper reports on the conceptual framework and design of a project intended to, (i) provide a comprehensive evaluation, in terms of cultural appropriateness, of existing teaching and learning systems provided on the Web by Australian universities; and, (ii) determine the requirements for an effective Web–based flexible learning system for culturally diverse students, for use by Australian and other universities.

Introduction

The Internet and especially the World Wide Web, is increasingly being used as a vehicle for flexible learning, where learning is seen to be free from geographical, time and participation restraints (Nguyen, Tan, & Kezunovic, 1996; Rossman, 1992). Indeed, large numbers of institutions of tertiary education are rapidly investing considerable resources and faith in the Web as a means of conveying both the administrative and the pedagogical materials for student learning. Australia figures prominently amongst the nations of the world which use distributed information systems such as the Web, to deliver education (Paulsen, 1992; Rudra, 1994). However, in many cases, it seems, paper–based information resources are simply being converted so that they can be accessed using the Web, without regard to appropriate design models and strategies for exploiting the Web as an instructional medium (Alexander, 1995; Reeves, 1996), particularly for students originating and studying with different cultural perspectives.

Distributed learning systems on the Web have the potential and often the intention of reaching greater numbers of culturally diverse students. The key to success in the use of the Web across cultural boundaries lies in the appropriate design of on–line educational environments (Harasim, 1995; Henderson, 1996). Our own recent research (Henderson, Patching, & Putt, 1996; Oliver, 1996; Wild, 1996; Wild & Omari, 1996) has demonstrated that the Web is almost always chosen as a delivery medium for instruction primarily for its ubiquity and insignificant costs; however, it is not chosen for its instructional effectiveness; nor is it chosen as a medium particularly suited to carrying a range of information types for culturally diverse learners. In this context, creating systems for teaching and learning on the Web may well work to limit efficacy in learning, despite the Web's ever–developing technical capacities to carry multimedia materials and information, and its growing provision for various levels of complexity in learner–material interactions. Indeed, while many learners might possess the basic information and navigational skills to contend with information access on the Web, instructional designers are yet to consider those aspects of this medium that determine its effectiveness for all learners, whatever their cultural characteristics. Henderson has determined, as late as 1996, 'the relationship between cultural context and instructional design has received little attention in the educational technology and instructional design literature' (Henderson, 1996 85). It seems apparent that the lack of research to target cultural issues in instructional design for distributed and interactive learning systems, such as those being placed on the Web by Australian and overseas universities in ever-increasing numbers, is even more noticeable and is likely to have serious consequences, particularly for students as well as for universities. Indeed, a recently published government funded report in the area of Internet use by universities, makes a clear recommendation that 'the university sector should be proactive in profiling and accommodating characteristics of user need that will require compatible network technologies' (Bruce, 1996 xi).

Flexible Learning and the Use of the Web

The concept of flexible learning has evolved alongside developments in new technologies in four main phases. The initial phases are considered by Nguyen, et al (1996), Peacock (1995), and others, to have focused on, (i) correspondence and radio broadcasts (directed at isolated learners in farming and mining communities); (ii) video and television broadcasts, which allowed greater approximation of the traditional classroom experience; and, (iii) computer conferencing, electronic mail and voice mail, which supported greater levels of synchronous and asynchronous communication across distance and time. A fourth phase is set to emerge in the later 1990's and will be focused on direct student access to computer based remote databases, hyper- and multi-media information and dial-up access to video materials. 'Students will (*in this fourth phase*) be in control of the time, place and pace of study' (Peacock, 1995 our italics), and will have direct access to an expanding and dynamic knowledge base and extensive communicative facilities. The Web is an emerging technology but already possesses the functionalities described in this fourth phase: the Web is placed to be the technology most likely to carry flexible learning into a new phase of development.

Of course, the move towards flexible learning for all students, on-campus and distance, is being driven not only by technological imperatives but also by economic and pedagogical ones. There is a declining ratio of academic staff to students; and students are increasingly being encouraged to invest greater independence in, and control over, their learning. Indeed, over the last 20 years or so there have been significant changes in policies, organisation, staffing, funding and management of universities in Australia, usually as a result of government directives and policies (Chalmers & Fuller, 1996). One consequence of these changes is that students are now a much more diverse group, particularly in cultural characteristics, and are more likely to study in mixed modes that are suited to flexible learning.

Educational Potential of the Web

The nature of the World Wide Web has attracted a great deal of rhetoric in favour of its potential to provide for a student-centred model of learning, where the learner is both intrinsically motivated and active in the learning environment (Becker & Dwyer, 1994). At first glance there is much in the Web that appeals to educators—for example, the hypermedia information structures in the Web allow for the chunking of information, a feature that, in light of information processing theories of working memory, might be seen to support the cognitive processing of knowledge (Biggs & Moore, 1993). There have also been suggestions that in providing for browsing and thematic exploration, the Web facilitates higher order cognitive processes, such as transfer and knowledge application (Jacobson & Spiro, 1995); whilst at a more conceptual level, there has always been a case made for hypertext mirroring the ways in which much of human thinking occurs, by association rather than linearly or procedurally (Burton, Moore, & Holmes, 1995; Bush, 1945; Minsky, 1975). Furthermore the Web, in terms of being a dynamic, extensive and extensible information base, provides for the ultimate in resource-rich learning.

It is important to remember that the Web as hypermedia or hypertext, is itself only a medium for conveying information or knowledge. Hypermedia does not possess a single or normative information structure—hypermedia documents are created to conform or fit to a structure, imposed by their instructional designers. At one extreme this structure might be highly ordered, supported by a constrained and sequential set of links; whilst at another extreme, the hypermedia may be nonsequential and supported only by referential links. In many cases, a coherent hypermedia document, such as a Web site, might comprise a mix of these structures. It is, then, the nature and application of these structures that determines the effectiveness of engagement with knowledge carried in the Web. Furthermore, to maximise engagement, the knowledge needs to conform to a structure that best fits or suits both the type of knowledge being conveyed as well as the learning preferences and requirements of diverse groups of learners.

However, there is no guidance, and virtually no empirical research, to help determine the most appropriate ways of using the Web to stimulate effective learning at tertiary level for all learners that are so targeted. It is apparent, however, that instructional design for Web–based learning systems cannot, and does not, exist outside of a consideration of cultural influences—both the cultural influences operating on the authors and instructional designers of Web–based learning materials, and similarly, those influences that impact on the interpretation of such materials by learners.

The Influence of Culture

Defining culture is a difficult proposition, and many different classifications exist in relation to national culture (Kluckhohn and Strondbeck, 1961; Roackeach, 1973; Hall, 1959, 1990; Hofstede, 1984; Hofstede and Bond, 1988). Perhaps the most pervasive view is that culture is a manifestation of ways in which an identifiable group adapts to its changing environment; that people belong to more than a single cultural group, embodying a subset rather than a totality of a culture's identifiable characteristics; and that they do not remain totally allegiant to their birth culture (Henderson, 1996; Scheel & Branch, 1993).

Whatever the definition, there appears to be consensus that culture must have a definite and very strong influence on the design and use of information, communication and learning systems, as well as on their management, despite the lack of identifiable research in these areas. In all areas of human activity, the behaviour of people is affected by the values and attitudes that they hold and the societal norms which surround them. When values are widely shared by a group of people, they are provided with a common mechanism by which they can share understandings and interpretations of their world, and establish what is important and clarify priorities. As nations develop and organisations become more technologically advanced and globally oriented, their culture changes and this, in turn, has an effect on individuals' attitudes and values (Adler, 1991). Culture, however, is more than just an abstraction, it also consists of a distinctive symbol system together with artefacts, that capture and codify the important and common experiences of a group. Distinctive significant symbolic meanings and values develop around information, its use and structuring in any cultural group. Also, at a practical level, when the act of instructional design translates this information into products or artefacts of learning, that artefact embodies cultural influences, such as the instructional designer's world view, their values, ideologies, culture, class and gender, and, their commitment to a particular design paradigm (Henderson, 1996).

These interacting cultural factors have a particular importance for the diffusion and efficacy in use, of information, communication and learning systems, such as the Web, and the products and materials of learning provided in those systems.

A Model for Investigation

We presently have a situation where cultural influences in distributed information, communication and learning systems, especially those centred in the Web, are present and are identifiable, but are largely created unknowingly. As a result, such systems probably work to the detriment of large groups of culturally diverse learners who cannot identify with the instructional designs in Web–based systems of teaching and learning, originating as they do, in single cultural identities. Given the instructional agendas currently being set by Australian universities for the present and future use of the Web, it is reasonable to suggest that there will be a mismatch between instructional intention and learning outcome. This mismatch will become more noticeable as Web–based flexible learning systems are increasingly put into place in the later 1990s and into the new millennium.

Henderson describes three existing instructional design paradigms in static instructional multimedia: (i) culturally unidimensional or exclusionary; (ii) inclusive; and, (iii) inverted (Henderson, 1996). In the first paradigm, cultural minority groups go unrepresented. Scheel & Branch (1993) have attempted to describe the reasons for this, and in doing so explain various manifestations of what Rattansi (1992) has termed, *deracialisation*. Deracialisation occurs when there is an unintentional or intentional exclusion or avoidance of, or insulation from, issues of appropriate cultural contextualisation in the production of multimedia learning materials. In the second paradigm, Henderson (1996) acknowledges the adoption of an inclusive or perspectives instructional design approach, where the instructional designer includes the social, cultural, economic and/or historical perspectives and/or contributions of minority groups. 'In this paradigm, instructional design is driven by social justice and equity issues, while instructional design solutions range from soft to hard multiculturalism' (Henderson, 1996 91), or what Scheel & Branch (1993) term 'mild to strong interventions' (p. 9). In a third paradigm, the instructional designer will attempt to approach the design task from the perspective of one or more minority cultures, that is, from an inverted curriculum or critical theory–postmodernist paradigm (Henderson, 1996 93).

Each of these instructional design paradigms have been determined by Henderson (1996), to be unsatisfactory in terms of providing culturally appropriate instruction in static (ie. CD–ROM) multimedia products. It is reasonable to hypothesise that these are the very paradigms that currently also dominate in distributed information, communication and learning systems, presently being provided on the Web by Australian universities for teaching and learning for culturally diverse students.

Research Plan

The lack of space in this paper, prevents us making a detailed report on the research methodology we are using in our present work, suffice to state that our selection has been guided by both Howe and Eisenhart (Howe & Eisenhart, 1990) and Reeves (Reeves, 1993), who argue that any methodology employed should be judged in terms of its success in investigating educational problems deemed important. Moreover, Salomon (1991) describes the contrast between analytic research that is focused on isolating effective instructional treatments and systemic research focused on understanding how instructional treatments work in practice. This suggests that analytic and systemic approaches are complementary: 'the analytic approach capitalises on precision while the systemic approach capitalises on authenticity' (Salomon, 1991 16). Both analytic and systemic methods are being used in this research programme. Also, the nature of learning based on the Web, with its high degree of individualisation, 'meshes precisely with the naturalistic assumption of individual constructions of reality' (Neuman, 1989 48). Indeed, specific strategies based on case study methods have been highlighted in our research programme, so we can elicit these individual constructions.

There are three phases planned in the research, corresponding to three temporal stages. The following broadly outlines each of these stages:

Phase I/Year One: The major focus for this phase is to identify the existing instructional design paradigms that exist in distributed information, communication and learning systems provided by Australian universities for culturally diverse groups of learners, both internal and external to Australia.
Phase II/Year Two: The major aim of this phase is to (a) create and implement of a number of Web sites based on a fourth instructional design paradigm centred in a view of multiple cultures (Henderson, 1996), and (b) implement a pilot study to test the paradigm's effectiveness.
Phase III/Year Three: This phase aims to investigate the degree of success in student learning that students of different cultural groups have as a result of using the Web-based materials designed in the paradigm of multiple cultures. The students (n>1500) chosen for investigation will include students enrolled internally and externally to at least two Australian universities, and who are therefore studying within and outside Australia.

They will also include students who can be, collectively, identified with a range of representative cultural groups. The study will incorporate quantitative and qualitative data collection instruments, such as:

- pre, post, and post-delay content questionnaires;
- pre and post attitudinal, anxiety, and usage questionnaires (King, Henderson, & Putt, 1997),to be incorporated within the Web site pages;
- a five category Likert questionnaire focussed upon the interface design features of the Web sites, to be incorporated within the Web site pages;
- case study of 12 ethnically diverse students;
- stimulated-recall audio-taped interviews with students at Australian and overseas study sites;
- observations of students at Australian and overseas study sites;
- other evaluative open-ended questionnaire interviews of students; and,
- tracking data.

One of the challenges to this research project will be the creation of data collection instruments that enable accurate identification of instructional design paradigms used in the distributed learning systems on the Web sites selected at Phase One. Such identification will be based upon how well the Web sites conform to critical identifiers of the three existing instructional design paradigms identified by Henderson (1996) that are hypothesised to exist in Web-based systems created for flexible learning. Currently, we are hypothesising that the instruments will involve checklists that include the following sorts of relevant instructional design elements seen to belong to the three cultural paradigms described in Henderson (1996):

- the underlying pedagogic philosophy of each Web course site;
- the Web course site's epistemology;
- each site's instructional sequencing (ie. is there a wholist (horizontal hypertext) or partist (linear hypertext) layout to the interface design of what is usually seen as the content menu page/s);
- the degree of in–built individual versus collaborative strategies;
- hypermedia navigation pathways that cater for individual learning styles;
- the ratio of Anglo/Western (ie. American, Canadian, British, Australian and New Zealand) active internet links, to non-Anglo/Western internet links included within each Web course site;
- count of key concepts and examples of, for instance, a single 'truth' or multiple theoretical perspectives;
- semantic chunking of text versus traditional linear structures;
- appropriate/inappropriate culturally contexted information; and,
- appropriate/inappropriate culturally contexted graphics, animation, video clips, sound, and colours.

It is also expected that such design elements will also, in some form, guide the design of a number of Web sites to be created, that are based on a fourth instructional design paradigm centred in a view of multiple cultures (Henderson, 1996).

Conclusion

The aim of this research project is to identify the nature and improve the efficacy of models of flexible, open and distance learning created in the World Wide Web by universities. The following hypotheses are central to our work:

- Existing cultural influences in instructional materials designed and delivered on the World Wide Web by Australian universities, and intended for use by culturally diverse students, are minimal and ineffective.
- The efficacy of learning based in the use of the World Wide Web for instructional purposes can be improved by the adoption of a culturally appropriate model of instructional design.
- Culture is a significant factor in determining the effectiveness of learning materials created in the World Wide Web and intended for use by culturally diverse students.

In testing these hypotheses we intend to provide the empirical research, to help determine the most appropriate ways of using the Web to stimulate effective learning at tertiary level for all learners, whatever their cultural heritage or perspectives.

References

Alexander, S. (1995). Teaching and learning on the World Wide Web. Paper presented at *AusWeb95: The First Australian World Wide Web Conference*, Southern Cross University, New South Wales, 30 April-2 May, 1995.

Becker, D., & Dwyer, M. (1994). Using hypermedia to provide learner control. *Journal of Educational Multimedia and Hypermedia*, *3*(2), 155-172.

Biggs, J. B., & Moore, P. J. (1993). *The process of learning* (Third ed.). Sydney: Prentice Hall.

Bruce, H. W. (1996). *Internet, AARNet and academic work: A longitudinal study* No. Department of Employment, Education, Training and Youth Affairs (DEETYA), Canberra. November, 1996.

Burton, J. K., Moore, M., & Holmes, G. A. (1995). Hypermedia concepts and research: An overview. *Computers in Human Behaviour*, *11*(3-4), 345-369.

Bush, V. (1945). As we may think. *The Atlantic Monthly*, *176*(1), 101-108.

Chalmers, D., & Fuller, R. (1996). *Teaching for learning at university* . London: Kogan Page.

Harasim, L. (1995). The virtual university: New approaches to higher education in the 21st century. Paper presented at *ASCILITE95. The 12th Annual Conference of the Australian Society for Computers in Learning in Tertiary Education* , Melbourne, 3-7 December, 1995.

Henderson, L. (1996). Instructional design of interactive multimedia: A cultural critique. *Educational Technology Research and Development*, *44*(4), 86-104.

Henderson, L., Patching, W., & Putt, I. (1996). Off-campus preservice teacher education via interactive multimedia technology: An indigenous cohort case study. In P. Carlson & F. Makedon (Eds.), *Interactive multimedia and hypermedia, 1996* (pp. 300-305). Charlottesville, VA: Association for the Advancement of Computers in Education.

Howe, K., & Eisenhart, M. (1990). Standards for qualitative (and quantitative) research: A prolegomenon. *Educational Researcher*, *19*(4), 2-9.

Jacobson, M., & Spiro, R. (1995). Hypertext learning environments, cognitive flexibility and the transfer of complex knowledge. *Journal of Educational Computing Research*, *12*(4), 301-333.

King, J., Henderson, L., & Putt, I. (1997). Measuring affective aspects of www and e-mail use in course delivery. In T. Muldner & T. C. Reeves (Eds.), *Proceedings of ED-MEDIA97 and ED-TELECOM 97: World Conferences on Educational Multimedia and Hypermedia and on Educational Telecommunications, 14-19 June, 1997* , (1) (pp. 595-600). University of Calgary, Canada: Association for the Advancement of Computing in Education (AACE).

Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision* New York: McGraw Hill.

Neuman, D. (1989). Naturalistic inquiry and computer-based instruction: Rationale, procedures and potential. *Educational Technology Research and Development*, *37*(3), 39-51.

Nguyen, A. T. A., Tan, W., & Kezunovic, L. (1996). Interactive multimedia on the World Wide Web: Implementation and implications for the tertiary education sector. In M. Nott (Ed.), *Proceedings of AusWeb96: The Second Australian World Wide Web Conference*. Gold Coast, Australia: Southern Cross University.

Oliver, R. (1996). Interactions in multimedia learning materials: The things that matter. In C. McBeath & R. Atkinson (Eds.), *Proceedings of The learning superhighway: New world, new worries? Third Interactive Multimedia Symposium* , (pp. 303-308). Perth, Western Australia: Promaco Conventions Ltd.

Paulsen, M. F. (1992). Computer mediated communications and distance education around the world. *From Bulletin Boards to Electronic Universities*, *7*, 31-37.

Peacock, K. A. (1995). *Distance education in a univesrity setting: connecting to the global classroom.* Available, http://www.utirc.utoronto.ca/Distance_Ed/disted-report.html [on-line]. University of Toronto, Canada, October, 1995.

Reeves, T. (1993). Pseudoscience in computer based instruction: the case of learner control research. *Journal of Computer Based Instruction*, *20*(2), 39-46.

Reeves, T. (1996). A model of the effective dimensions of interactive learning on the world wide web. In P. Carlson & F. Makedon (Eds.), *Interactive multimedia and hypermedia, 1996* Charlottesville, VA: AACE.

Rossman, P. (1992). *The emerging worldwide electronic university: Information age global higher education* . Westport: Greenwood Press.

Rudra, A. (1994). AARNet and ERNET: An era for fostering international academic research. In N. F. Ellerton (Ed.), *Proceedings of International networking: Education, training and change*, (pp. 104-109). Perth, Western Australia: Edith Cowan University.

Salomon, G. (1991). Transcending the qualitative–quantitative debate: The analytic and systemic approaches to educational research. *Educational Researcher*, *20*(6), 10-18.

Scheel, N., & Branch, R. (1993). The role of conversation and culture in the systematic design of instruction. *Educational Technology*, *3*(3), 5-14.

Wild, M. (1996). Designing appropriate learning environments in instructional multimedia: The role of theory. Paper presented at *Higher Education Research and Development Society of Australasia (HERDSA) Annual Conference: Different approaches—Theory and practice in Higher Education* , Perth, Western Australia, 8-12 July 1996.

Wild, M., & Omari, A. (1996). Developing educational content for the Web: Issues and ideas. In R. Debreceny & A. Ellis (Eds.), *Proceedings of AusWeb96: The second Australian WorldWideWeb Conference. (Also available at, http://www.scu.edu.au/ausweb96/)*, (pp. 353-355). Gold Coast, Australia: Southern Cross University.

PANELS

WIRED COMMUNITIES – CAN WE CREATE AFFINITY THROUGH ELECTRONIC VICINITY? : A PANEL

Ron Riesenbach, President Telepresence Systems, Inc. Toronto ron@telepres.com
Tom Jurenka, Director, Disus, Toronto tom@disus.com
Dr. Gale Moore, Associate Director, Knowledge Media Design Institute, University of Toronto, Toronto
gmoore@dgp.utoronto.ca

The wired community is a one of the hot applications emerging out of the Internet today. It seems that every large telecommunications, computing and media firm in the world is involved in at least one big wired community project. Some are turning out to be successful, others are quietly shutdown after having turned into expensive disasters. Something is wrong.

What is becoming clear from these trials is that while the technologies are complex and expensive, the sociology of human-to-human and human-to-machine interaction is even more difficult. It is our observation that technology doesn't create communities, people do. However, we do believe that technology can be used to support and extend communities if it is done right. The root of the problem is understanding the nature of the concept we call "community".

This panel includes individuals who have been instrumental in the development and operation of a number of state-of-the-art electronic communities in Canada. They will share their experiences and lessons learned, and invite discussion on the issues raised.

HOW TO CREATE A WIRED COMMUNITY DESPITE THE LATEST TECHNOLOGY

Ron Riesenbach

It is my observation that there is a common misconception in the Information Highway business that more equals better. "More bandwidth, more pixels, more frames per second..." yell the high-tech hucksters as if this were the answer to all our problems. The premise behind their strategy is that geographically dispersed people can be thrown together in a maelstrom of technical virtuosity, and that communities will emerge at the other end. They then hope that while "the community" sits in stunned amazement in the glow of their 3D VRML electronic avatar chat room, that they can sell them a stream of pre-digested "content" for their "interactive" pleasure.

It is encouraging that this view of the world is not universal. In fact, findings from several wired-community initiatives that I have been involved with are now showing that this techno-centric approach is as misguided as we expected. In this presentation I will draw examples from specific initiatives and demonstrate some of the technologies we've developed. I'll show how the use of user-centred application design and deployment strategies have contributed to the success of electronic communities – despite the latest technology.

SOUNDS GREAT - GOTTA GO !

Tom Jurenka

Widely available interactive networks such as the Internet hold great promise for community oriented information services. Closer integration and freer flow of information between parents and schools, citizens and municipal organizations, voters and policy makers, to name just a few, promise to lower the barriers between an individual and those parts of the world important to him or her.

However, this rosy picture can only be realized if a large portion of the population participates in and contributes to the growth and development of interactive community based content and services. Without such participation, interactive networks will become another vehicle for pre-digested, anaesthetized information of the type provided by the majority of media outlets today. The challenge is not just to interest users in content creation, but to keep them interested and engaged, to provide appropriate tools, and to help them make time for such activities in the middle of busy lives.

For the last year, as Trial Manager of Intercom Ontario, I have been working with community resources such as schools, health providers, and public organizations in Newmarket Ontario, site of the Intercom Ontario broadband trial. I'll discuss my perspective and my experiences in attempting to bootstrap an interactive community, and suggest how such efforts might be organized elsewhere.

"DO FENCE ME IN" - BOUNDARIES AND BORDERS IN CYBERSPACE

Dr. Gale Moore

Cyberspace has been described as the new frontier and while we busily shed old limits as we meet, play, work increasingly independent of time and space it's easy to forget the technologically mediated nature of our activities, and to assume that connectivity can be equated with community.

By the early 1990s computing had become increasingly social and by 1996 the concept of community had become the concept de l'annee. Corporations, government reports, conferences, even technical conferences used the term freely, in part impelled by the idea that the Internet and Web were bringing us together in new ways in what have been called "virtual communities" or "communities of interest". While this has drawn attention to the social dimension of networked environments, I suggest that it has also led to an impoverished view of "community" and of the social potential of community development in Cyberspace.

Community is a powerful concept and understood sociologically to be something richer and more complex than the long-term ongoing dialogue that characterizes many virtual communities. Community suggests such things as membership, commitment, shared values, and reciprocity. Communities need borders and boundaries – not in a negative sense to isolate or protect themselves, but in the positive sense of "knowing where you are", and the sense of belonging and trust that fosters creativity and the sharing of ideas and plans. The key is that the boundaries be semi-permeable – providing private places and public spaces within the community as well as access and exchange with those outside.

I will demonstrate through the example of the Virtual Sandbox – a community development environment – one way in which these ideas are being explored. Taking a human-centred approach to design we work in partnership with members of existing communities to develop an environment of services and applications that accommodates both public and private activities and that are readily tailorable by each community. Our goal is to support and enhance the ability of community members to work and play together in more meaningful ways than are currently possible and to explore how communities migrate and reproduce themselves in networked environments.

POSTER/ DEMONSTRATION SESSIONS

A JAVA Toolkit for Representing a 3D Environment

Marco Aiello
Università di Roma *La Sapienza*
Via Vetulonia, 38A 00183 Roma (Italy) tel +39-6-70475767
aiellom@acm.org

Pierluca De Maria
Università di Roma *La Sapienza*
Via Iris Versari, 218 00128 Roma (Italy) tel +39-6-50795763
p.demaria@usa.net

Cristiano De Mei
Università di Roma *La Sapienza*
Largo Beltramelli, 36 00157 Roma (Italy) tel +39-6-41732181
mc7655@mclink.it

Acknowledgements

The authors are grateful to Professor Marco Schaerf for his technical advice.

The Project

The main aim of the present work is the representation and visualization of 3D environments on the Web enabling high user interactivity. Technology available until now includes VRML, QuickTime VR, ActiveX and Java; but the first two lack interactivity, the last two lack 3D authoring tools. We have chosen Java and we have built a Java library to render 3D objects, taking advantage of Java's interface components, event generation and event handling. We have employed the above library to create a virtual world of colored blocks as perceived when navigated by a user controlled bodyless robot.

The idea is that of a moving robot in a randomly generated world. The user sees on the screen the moves the robot performs and also a 2D map indicating the position of the robot with respect to the world. The robot is controlled by the user via the keyboard or via the buttons on the applet's interface.

The demo applet can be viewed at http://www.d is.uniroma1.it/~aiellom/webnet97/robot2.html, the full documentation at http://www.d is.uniroma1.it/~aiellom/webnet97/tesina.html. Please feel free to contact any of the authors for any further information.

Utilizing Technology to Enable Practitioners to Serve as Mentors for Preservice Teachers: Strategies for Research

Gregory F. Aloia, Ph.D.
Professor
Department of Special Education
College of Education
Illinois State University
Normal, IL 61790, USA
gfaloia@rs6000.cmp.ilstu.edu

Ramesh Chaudhari, Ph.D.
Management Systems Coordinator
Information Systems
College of Education
Illinois State University
Normal, IL 61790, USA
ramesh@rs6000.cmp.ilstu.edu

Jeffrey P. Bakken, Ph.D.
Assistant Professor
Illinois State University
College of Education, SED Box 5910
Normal, IL 61790, USA
jpbakke@rs6000.cmp.ilstu.edu

John Beemsterboer
Graduate Student
Department of Special Education
College of Education
Illinois State University
Normal, IL 61790, USA
jjbeems@rs6000.cmp.ilstu.edu

The induction into the profession of teaching can begin at the preservice level by establishing an electronic mentoring relationship (via the use of the Internet and email) between practitioners in the field and preservice teachers who are enrolled in an undergraduate training program. An electronic mentoring relationship would not only increase the usage of technology among teachers and serve as a part of "pre-induction" experiences for students into the profession, but it would also result in an increased partnership between faculty in the colleges of education and teachers in the field. It would initiate the merger of technology, professional partnerships and induction.

Key factors that influence the success of a mentor/protégé relationship are addressed as well as strategies for utilizing e-mail and the Internet as a means of supporting preservice education majors. Several specific suggestions for research are provided, along with critical questions that should be examined.

Using Evaluation to Monitor and Improve the Early Implementation of Online University Teaching: an Australian Example

Pamela F. Andrew and John M. Owen
Faculty of Education
The University of Melbourne
Australia

p.andrew@edfac.unimelb.edu.au
j.owen@edfac.unimelb.edu.au

We know from the literature that the adoption and use of a major innovation such as online distance education cannot be regarded as a *fait accompli*. The introduction of a major new way of providing teaching and learning through the WorldWideWeb involves radical changes in the way teachers, learners and support staff behave. These changes cannot be put in place overnight. The innovation literature shows how important it is for changes to be phased in over time. In cases where little prior information from other sources is available about the innovation, there is a role for evaluation to provide information which can be used to make decisions related to the change process. We report here on a case study which puts into practice these principles, in particular, we examine the role of evaluation in the development of courses in the areas of Program Evaluation and Early Childhood Studies.

Enabling Distance Education over the World Wide Web

I. Antoniou[1], E-mail: antoniou@ceid.upatras.gr

C. Bouras[1,2], E-mail: bouras@cti.gr

P. Lampsas[1,2], E-mail: lampsas@cti.gr

P. Spirakis[1,2], E-mail: spirakis@cti.gr

[1] Computer Engineering and Informatics Department
University of Patras, 26500 Rion, Patras Greece
[2] Computer Technology Institute, Kolokotroni 3, 26221 Patras, Greece

The recent advancements in the field of the information and communication technology resulted in a revolution comparable to the industrial revolution. The basis of this revolution is the information and the value it has as the pure expression of human knowledge. The technological advancements offer us the ability to process, store, retrieve and transmit information in multiple formats (text, sound, image, video) independently of time, volume and distance. In this paper we present the technology that is required and an architecture for the realization of distance education over the World Wide Web. We also address some basic aspects of the users' needs that every software tool for distance education should meet.

Growing Usage of Internet Services: A View From A Developing Country

Cengiz S. Askun
Department of Instructional Systems Technology
Indiana University, Bloomington, IN 47405
e-mail:cengiz@indiana.edu

Kursat Cagiltay
Department of Instructional Systems Technology
Indiana University, Bloomington, IN 47405
e-mail: kursat@indiana.edu

Wide Area Networks in Turkey have been first organized in 1986 as an extension of EARN. Internet connection realized in the beginning of 1993 was a significant step forward. It has become wide-spread with a tremendous rate of growth among various sectors in last years.

This poster reports recent and current use of the Internet and WWW based services in a developing country, Turkey. Some common problems most of the developing countries are facing, and some of the experiences are presented.

This poster especially focuses on two different avenues: the use of the Internet in educational institutions and commercial use of the Internet in Turkey. A statistical information will be provided along with a comparative analysis of the past years in terms of the commercial and educational usage of the Internet. Some comparisons will be made with the developed countries. On the other hand, some problems (cultural, social, technical, etc.) of the Internet in developing countries will also be discussed in this poster.

The World Wide Web: A new approach to world wide supply of epidemiological data on diabetes mellitus

Thomas Baehring (1), Hardy Schulze (1), Stefan R. Bornstein (1), Werner A. Scherbaum (2)

(1) Department of Internal Medicine, University Hospital of Leipzig, Germany
(2) Diabetes Research Institute c/o University of Düsseldorf, Germany

Contact:
ThomasBaehring, PhD
University Hospital of Leipzig
Department of Internal Medicine
Ph-Rosenthal-Str. 27
04103 Leipzig, Germany
Tel: (+49 341) 97 13 202
Fax: (+49 341) 97 13 259
E-Mail: baet@server3.medizin.uni-leipzig.de

Diabetes mellitis is a major health problem with a growing population. Presently there are approximately 100 million diabetics worldwide. Due to changing lifestyles and an increase in life expectancy, the developing nations are now increasingly at-risk along with the industrialized nations. Approximately 90% of diabetics are Type-2. Often Type-2 diabetes goes undetected in the early stages because the symptoms seldom manifest themselves. A diagnosis of diabetes is then made too late, when complications arise (such as stroke, heart attack, damage to the small and large blood pathways, eye disease) which require medical attention.

In order to prevent diabetes in a timely manner a certain diabetes Type-2 educational program and a successful disease-management plan will be necessary. It should provide epidemiological data with individual, genetic and social risk factors, all of which should be both simple to retrieve and internationally applicable. This poses a problem for the present, for the provision and analysis of At-Risk tests on such a large scale is extremely expensive. In addition, the data-retrieval systems vary regionally, so that an international applicability of test results until now is practically impossible.

In light of this, we have prepared an interactive CGI-Test form on the World Wide Web with questions regarding the common diabetes risk-factors which have world-wide applicability. The corresponding characteristics (age, height, weight, sex, known diabetic history, movement restriction, familial predisposition to diabetes, macrosomie) are then entered by the web-site user directly into the PC and sent via Internet to our WWW-Server. While the user is still online he/she receives an evaluation of his/her personal risk for becoming diabetic, and the data are saved confidentially on the WWW-Server, including only the server address, date and time. The publication of the WWW-Address of the Risk-Test travels to all German media, as well as worldwide through mailings over the Internet (mailing lists, news groups, personal emails).

Up to this point 1,400 responses worldwide have been registered, with an increasing frequency (60.9% from Germany; 9.4% overall in Europe; 18.9% from North America). The respondants were on the average 38.3+/- 14.0 years of age (men 39.8; women 35.9) with an age range from 10 to 83 years. 76.0% of the web-users tested negative. 66.8% of the respondants had restricted movement; 40.4% had a family history of diabetes; and 19.6% of the women who responded had given birth to an overweight baby (macrosomie). Surprisingly, for 45.3% of the test-users an increased risk of diabetes was indicated through the classification algorithms. Analysis of the entered data in regard to a country-specific spread of risk-factors were possible with 51.2% of the data through Server-IP-Addresses.

The World Wide Web proves itself to be a time-saving and cost-efficient technical platform which gathers epidemiological data regarding the risk-factors of diabetes with a world-wide applicability, thereby reaching a targeted audience that encompasses 45.3% (an above-average number) of people with a growing risk of diabetes. This clearly proves a positive correlation between the growing number of Type-2 diabetics and Internet-Users. Because the posed questions are clearly answerable, the credibility of the test results is very high. False entries can be identified and eliminated by analysis of the database.

Table 1

Structure of the data base on our W3 server for collection of statistical and epidemiological data from the user (indicated by an individual example).

data fields	sample record
host domain name	sunlight.ccs.yorku.ca
host IP-address	130.63.236.85
date (day)	10
date (month)	2
time (hour)	2
time (minute)	8
diabetic	no
sex	female
age [years]	56
height [cm]	168
weight [kg]	58
Body-Mass-Index [kg/m^2]	20.5
little physical activity	no
family history of diabetes	no
macrosomic infant	no
risk for diabetes	normal

Changing Ages: Transforming Paradigms, Policy andPedagogical Practice

Ivan W. Banks, Ed.D   Ruth R. Searcy, Ed D   Mike Omoregie, EdD
Jackson State University, School of Education

Education is generally regarded as the institution mostresponsible for providing "survival skills" needed to empower individuals to function effectively within thesocio-economic system of a nation. In American society, these skills are more or defined by theprivate sector which serves as the major source for employment relied upon by most individuals to sustaina livelihood for themselves and their families. As such, the schooling process, thoughvariously defined, is expected to provide common patterns of experiences and knowledge consideredessential for promoting economic growth.

As we approach the 21st century, technology has become thedriving force in the delivery of instruction to today's youth. Since the birth of microcomputers, theeducation community has recognized that teachers' training would be essential to the successfulintegration of technology in classroom instruction. While much has changed over the years, the needfor teachers' support and training has not. However, the importance of training teachers andadministrators remains the key to successful implementation of technology in the classroom.

A new paradigm termed techo-literacy is a means of fosteringthe development of the skills in literacy, numeracy, the humanities and technologies that are necessaryto negotiate economic self-sufficiency in this society. It is the onlyreliable way of combating the social determinism that now condemns the poor or African Americans to remain in social and educationalconditions of inequality. Techno-Literacy suggests we, as educators, have a responsibility to makeschools accountable to the needs of all children. They must be given the opportunity to learn andutilize skills for functioning in this highly technological society.

In conclusion, this paper discusses the changing ages withemphasis on transforming paradigsms, policy and pedagogical practice. Research suggests thatanalyzing the historical perspective of the Industrial Revolution reinforces the need fortechnological innovations and the usage of techo-literacy.

Real-Time Streaming Video for Ethernet-based Intranets

Jerzy A. Barchanski
Brock University
Department of Computer Science
St. Catharines, Ontario,
Canada, L2S 3A1
jbarchan@sandcastle.cosc.brocku.ca

We discuss first some measurements of Ethernet traffic which indicate that it can best be described statistically using the notion of self similarity It was observed not only in the Ethernet traffic patterns but as well in the sizes of the files residing in the file servers and variability of compressed digital video frame sizes.

We describe then our simulation model developed in COMNET III environment and experiments we have carried on it. These experiments let us to specify the best communications protocol, the maximal number of clients, the most appropriate message length and transmission speed for transfer and real-time display of digital video on Ethernets. We formulate then some requirements which should be satisfied by a streaming video system for Ethernet-based Intranets.

In conclusion we analyze and compare the suitability of a number of products developed for streaming video on the Internet from the viewpoint of these requirements and suggest some necessary modifications.

The Noteless Classroom

Robert Beard
Department of Modern Languages, Literatures and Linguistics
Bucknell University
USA

Many on-line syllabuses include a set of the lecturor's notes. Research, however, has shown that note-taking is an important part of the learning process. The projection of notes during lectures should offset the disadvantage of losing this aspect of the learning process. Relieving students of any stenographic responsibility should allow them to follow lectures and discussions more closely, without losing track and equalize the advantage of good note-takers over others. Reducing the instructor's use of the blackboard should also contribute to the cohesion of the dialog between the instructor and the class. This preliminary study demonstrated that most students (9 of 13) continue taking notes to assist in first-learning and to add detail to the on-line notes. Grades and student comments indicate that additional questions and discussion may contribute to improved learning.

Information Technology in the Second/Foreign Language Classroom: A Unit on Food

Mercè Bernaus
Departament d'Ensenyament
Barcelona (Spain)
mbernaus@pie.xtec.es

The use of the "information technology" in the second/foreign language classroom helps to:

- motivate students to look for information
- motivate students to take part in international projects
- establish contacts with other schools
- use the language in real situations: discussing with students from other schools through Internet, creating their own new pages, sending messages through e-mail...
- increase students' confidence and self-esteem
- develop new strategies in the process of language learning
- make use of different kinds of information resources
- ask for information to solve problems
- exchange information
- correct and/or enrich a written/oral text
- improve student's aural competence
- correct the oral production by themselves
- interact with various audiences
- encourage the exchange of ideas
- increase co-operation
- increase students' creativity

To make teachers aware of the advantages of the use of the "information technology" we provided them with materials ready to be used in the classroom. The main topic was FOOD, the technology resources needed were:

- cassette recorder
- word processor
- CD-ROM
- Internet

The learning tasks were the following:

- Listening and speaking: The recipe surprise.
- Reading for specific information and developing information strategies: Scavenger Hunt- Mama's Cucina.
- Co-operative work: stating hypothesis before searching information, collecting data, selecting information, agreeing on working process, reaching conclusions, presenting information. Linking FOOD with other areas of knowledge: Science (Nutrition), Art, Countries, The Third World, Nature. (Internet- WWW-, CD-ROM, Word Processor).
- Exchanging information with other groups outside school (Word Processor, Internet, e-mail). Eating customs contributions, e-mail messages.
- FINAL TASK: A REAL FOOD PARTY.

Large Scale Remote Graduate Instruction in Beam Physics

Martin Berz (berz@pa.msu.edu), Béla Erdélyi (erdelyib@pilot.msu.edu),
Jens Hoefkens (hoefkens@pilot.msu.edu)
Department of Physics and Astronomy and National Superconducting
Cyclotron Laboratory, Michigan State University, East Lansing
MI 48824, USA

Abstract

This article describes an Internet-based course on Beam Physics being offered by the Department of Physics at Michigan State University in the spring term of 1997. This course is part of the MSU Virtual University program and had about 100 registered participants at approximately 25 different sites all over the world. For the purposes of this course we are using ISDN-based and Internet-based videoconferencing tools, Internet-based transmission of audio and video recordings of the lectures, an interactive Internet-based homework system with on-line grading, and we provide the participants with downloadable lecture notes in a variety of formats.

Beam Physics is one of the more recent subfields of physics, and is connected to the understanding and development of particle accelerators. Because accelerator laboratories are usually not directly connected to university environments, the proper training of beam physicists at these sites often does not happen as naturally as in other fields. The availability of video conferencing and other Internet-based tools offers students and employees an option of increasing or refreshing their knowledge of Beam Physics. This approach provides an efficient and inexpensive mechanism to learn in a systematic fashion and offers the opportunity to earn university credit without leaving the workplace.

Introduction

Beam Physics became an important subfield of physics which incorporates many practical applications and challenging theoretical aspects. It is connected to the probing of the fundamental properties of nature and the search for new physics through high energy accelerator experiments, which represents the largest scientific experiments, to elucidating the structure of huge biological molecules trough mass spectrometers, to visualize tiny surface details trough electron microscopes, to fabricate computer chips trough micro beam litography, to build CRT's for TV sets, to separate isotopes, to measure exotic nuclei, and a variety of other techniques.

New theoretical methods developed for beam physics are at the forefront of physics, and they not just facilitate the understanding of current scientific instruments and provide solutions for future ones, but they are of academic interest by themselves, and in many cases their applicability goes far beyond the domain of beam physics. These are some of the reasons why a few years ago the Division of Physics of Beams was created within the American Physical Society. Nevertheless, there are many inconveniences and obstacles related to Beam Physics education. Due to the nature of this field, the highly trained instructors and specialists are spread over universities and major research laboratories. In the United States, only a few universities provide Beam Physics curricula, and one of them is Michigan State University. A significant amount of instruction is provided by the U.S. Particle Accelerator School via biannual two-week block courses at various locations in the USA as well as some other similar institutions.

From the above reasoning, the *necessity* to organize such a tele-course became clear; the goal is to provides large scale access to Beam Physics instruction, even internationally. Beside that, there are many further advantages: the convenience of taking such a course without leaving the workplace or school, reduced cost, flexible scheduling, the ability to gather recognized specialists to give guest lectures on their topic of expertise, etc. This paper describes what has been accomplished in this direction, and we think it provides the first step toward the future's virtual classroom.

The following table gives an overview of the participating sites, locations and number of participants per site.

Site Name	Country	Number of participants
Argonne National Lab.	USA	21
Beijing University	CHINA	1
Brookhaven National Lab.	USA	8
Calcutta University	INDIA	3
T. Jefferson National Lab.	USA	11
Cornell University	USA	1
DESY	GERMANY	1
Dubna Laboratory	RUSSIA	2
Fermi National Accel. Lab.	USA	7
Kansas State University	USA	2
KVI	NETHERLAND	4
Los Alamos National Lab.	USA	1
LawrenceBerkeley Nat. Lab.	USA	6
Lawrence Livermore Nat. Lab	USA	2
Mississippi State University	USA	1
Michigan State University	USA	11
Sandia National Lab.	USA	1
Stanford Linear Accel. Center	USA	1
St. Petersburg State University	RUSSIA	5
Stony Brook Laboratory	USA	1
TRIUMF	CANADA	3
University of Chicago	USA	1
Univ. Of Illinois, Chicago	USA	1
Universite Laval	CANADA	1
University of Helsinki	FINLAND	1

Table 1: Participating Sites

The subsequent sections will give some information about the importance of local contact persons, the technical aspects, used equipment and the Internet based homework assignments we have used. Finally we discuss the difficulties we have encountered and summarize our conclusions and future plans.

Technical Details

The method of attending the lectures allows a classification of the participants of our course in 4 major groups, with a subdivision of one of them in two subgroups.

- PictureTel User
    - Online User
    - Off-line User

BEST COPY AVAILABLE

CU-SeeMe User
Real Audio User
Videotapes User

The chart [Fig. 1] at the end of the paragraph gives a schematic overview of the equipment used and the flow of information to the participants for the various groups.

PictureTel

The major lab sites in the US are equipped with ISDN-based teleconferencing tools like PictureTel [PictureTel, 1997]. These sites use this equipment to participate in the lectures. Based on ISDN, the integrated digital data service, it offers higher bandwidths than ordinary phone lines and wide availability. While these solutions by themselves offer only point-to-point connections, Lawrence Livermore National Laboratory in California offers a service that allows more than two sites to connect to one video conference [VCS, 1997]. We are using this service (bridge) to serve the PictureTel based participants. This service is provided to institutions supported by the US Department of Energy for free; it is also available commercially at a rather affordable rate of about $40/hour. But the equipment for the system lies in a price range of about $20,000 up to $100,000, which naturally restricts the users of this system to participants located at the major lab sites in the US and Europe. While this system in principle has no restrictions on the locations (in fact we successfully transmitted lectures from Hamburg, Germany), all but the Hamburg site are located in the US.

Since the participants in this system actually see the other participants on a regular TV screen and the video signal is highly compressed, special arrangements have to be made in order to transmit a readable picture of the lecture notes. A designated document camera is used and the text has to be slightly bigger than usual.

Due to the time shift in the US and our lecture schedule (we started at 9:45AM EST), the lecture took place rather early in the morning at the West Coast. The participants at these places recorded the lecture with a normal VCR and watched the lecture later off-line (PictureTel Off-line User Group). The members of the PictureTel Online Group did attend the lectures live and online.

CU-SeeMe

The second group of our participants were the CU-SeeMe users. CU-SeeMe is a videoconferencing technology developed at Cornell University that uses regular TCP/IP to transmit highly compressed video and audio data over the Internet [CU-SeeMe, 1997]. The major advantage of this technology is that the necessary soft- and hardware is free or rather inexpensive. What all the participants of this group do actually need is a personal computer (either Windows based or a Mac), a working network connection (fast modem connections are in fact sufficient) and a sound card. Altogether these equipment needs can almost be considered standard for modern personal computers.

In addition, the participants do need the CU-SeeMe client software. There are two choices to obtain the software: it can either be purchased from a commercial vendor, or one can use the freeware client from Cornell University. Although the commercial product has some enhancements over the free version, we do not use it in order to maintain compatibility with the participants that do use the free version. Even the commercial product is at a price of $80 and hence rather affordable. On the server side a Windows-based PC with a frame grabber card is needed, at a cost of currently in the range of $250. It is not necessary to use commercial software in order to transmit the signals. In order to operate the CU-SeeMe part of the video conference, the participants log in to the *deflector*, which - as its name suggests- mirrors the incoming videoconference signal to the participants.

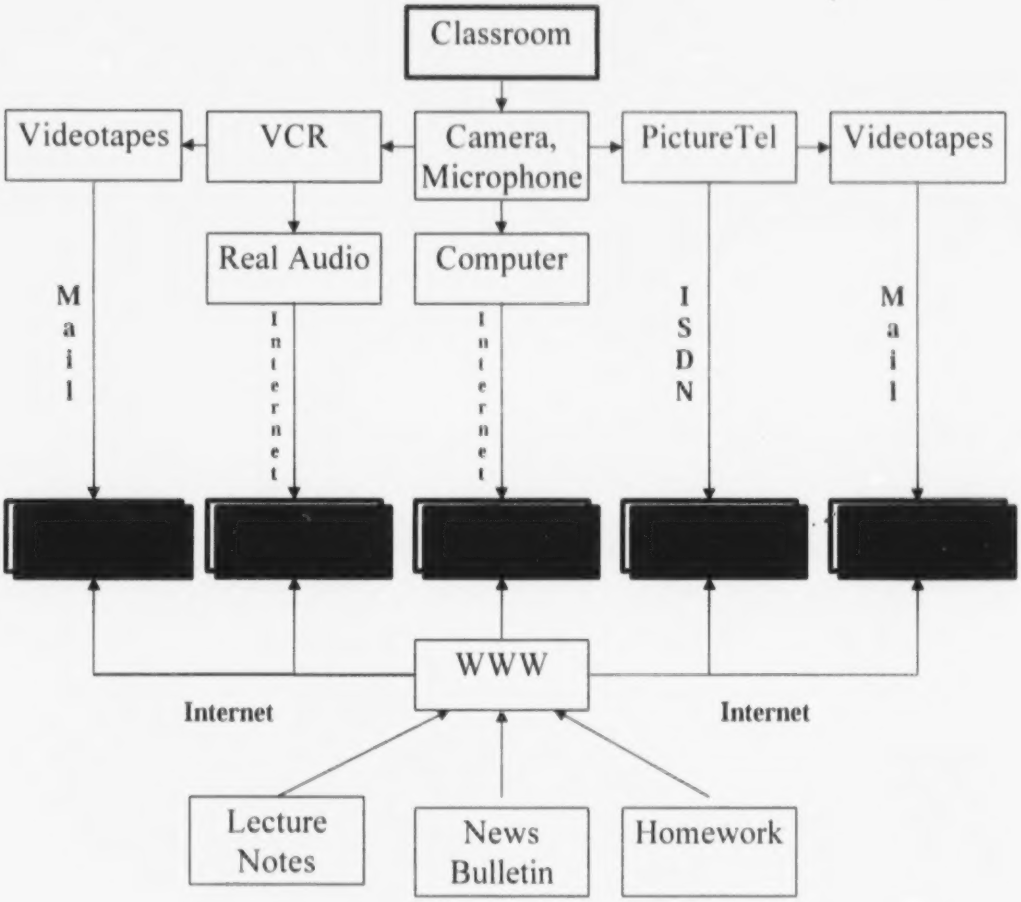

Classroom
Videotapes
VCR
Camera, Microphone
PictureTel
Videotapes
Real Audio
Computer
Mail
Internet
Internet
ISDN
Mail
WWW
Internet
Internet
Lecture Notes
News Bulletin
Homework

Figure 1: Information flow of the course

Real Audio, Video Tapes

In order to provide participants that can not attend the lecture live with as much information as possible, we make an audio recording of the lectures available on the Internet. These audio files are encoded from the video tapes that we have recorded during the lecture. The file format is the Internet standard Real Audio from Progressive Networks [Real Audio, 1997]. The participants can download these files a few hours after the lecture and listen to them while they follow the lecture notes that are available on the web. Other participants are provided with copies of these video tapes and get them sent by mail.

Both of these methods of participating are suitable for participants, that take the course on a standalone basis, cannot afford the PictureTel equipment, have very slow Internet connections to US sites, cannot take the lecture due to time shift problems, or who just missed the class for whatever reason. Due to recent developments by Progressive Networks towards a high-compression video standard for off-line viewing, we hope that in the near future it will be possible to provide the off-line participants not only with audio data but also video recordings of the lecture. However, for users with low bandwidth connections, the resulting files may become too large, considering that the audio files alone require about eight megabytes per lecture.

Local Contact Persons

It became clear early on that some participants would not feel comfortable with an Internet-based course that lacks the personal contact with the main instructor. Therefore it was of prime importance to us to establish close connections to qualified local contact persons that can help the students with questions and can provide personal contact. In some places these instructors even formed small groups with their students that were working together through the course material in local lectures. To provide further assistance for the participants, we established local office hours. During a fixed time slot we were offering the students the possibility to reach us by phone, fax, electronic mail and CU-SeeMe to ask questions regarding the lectures and the homework problems. We think that the general aspect of accessibility of the main instructor and the availability of help closeby cannot be stressed enough in the preparation of any distance education course.

Homework

For homework assignments, we are using CAPA [CAPA, 1997], [E. Kashy, 1994; E. Kashy, 1995], which is a software tool to implement a Computer-Assisted Personalized Approach for homework assignments, quizzes, and examinations. It was developed in such a way that it provides each student with a personalized assignment or examination with both quantitative and conceptual qualitative questions. With CAPA, an instructor can create problem sets which include pictures, graphics, tables, etc., with variables that can be randomized and modified for each student. Students input the solutions via a standard web browser, are given instant feedback and relevant hints, and may correct errors without penalty prior to the assignment due date. The system records the students' participation and performance in assignments, quizzes and examinations; and records are available on-line to both the instructor and the individual student.

CAPA was developed through a collaborative effort of the Physics-Astronomy, Computer Science and Chemistry Departments at Michigan State University, and the current version 4.5 became available April 15, 1997. However, more advanced features are needed in order to assign homework problems suitable to graduate students. Currently, the basic types of problems include numerical, multiple choice, matching, and true-false types, but there are no provisions for analytical derivations. At present, these type of problems can be hand-graded, and then the earned points added manually to the Grader module of CAPA. However, even in this case, the instructor saves some time compared to the completely hand-graded style. Of course, CAPA becomes clearly time saving for big classes. Future developments will provide support for other type of problems as well.

Publishing of Scientific Texts on the WWW

While we were preparing this course we were confronted with the need for evaluating the currently available mechanisms to publish scientific text on the WWW. One major aspect was the need to have a tool that allows the transformation of texts that are written in LaTeX, the main text editing software supporting complicated mathematical expressions, to HTML documents. Since most scientific texts containing extensive mathematics are written in LaTeX format, it seemed natural to us that there should be a way to publish these documents on the web. Two major aspects have been important for us. First of all the transformation should be as easy and as compatible with any computer platform as possible, and second of all the result should be esthetically pleasing.

While there is a tool that transforms LaTeX input files to HTML documents (LaTeX to HTML) the results are unfortunately often far from being pleasing, since every single equation is transformed into a separate GIF file. The resulting documents loose nearly all the nice formatting typical of LaTeX, and due to the literally hundreds of GIF files the loading times of the resulting HTML documents are unacceptable. Besides this tool there are some other solutions for LaTeX publishing on the web (TeX Explorer from IBM - a Netscape plug-in for Windows [IBM Alpha Works, 1997], Scientific Notebook - a proprietary browser from TCI Soft [TCI Software, 1997]). But all of them have the shortcoming that the software solutions are restricted to certain browsers and/or certain platforms. Since we had to maintain compatibility to all possible platforms, these solutions ruled themselves out.

There are also Java applications that can display mathematical equations in a very nive way. They are based on the above mentioned LaTeX typesetting system and the utilization of these tools is an ongoing effort in our preparation of future online courses. But based on current experiences, it seems to us that the publishing in PostScript format, with all its shortcomings, will remain the standard for distributing scientific texts containing mathematics in the near future. Another promising approach towards this problem seems to be the future development of the Portable Document Format (PDF) by Adobe [Adobe, 1996].

Conclusion and Future Developments

Overall we rate the Internet- and video conferencing-based course in Beam Theory we have given in the spring term of 1997 at Michigan State University as a full success. Not only have we reached a wide audience, but we also gained experience that we will use in future projects on distance education. Certainly this method of remote graduate instruction has its place in modern education and will become even more important in the next years, while improvement of the used technologies and methods will be an ongoing effort.

Since all the material is available in a web browser readable format, it is a natural extension to produce a CD-ROM out of the course material. This would allow students to take the course as it fits their needs and independent of the curriculum at Michigan State University that offers this course only in odd-numbered years. Given the interactive homework approach via CAPA, it is even possible to award credit to each individual who has at his own pace completed a full set of CAPA problems. This CD-ROM could be viewed with any web browser and could even contain the necessary additional programs (like the Real Audio player and GhostView).

Last but not least, we are already thinking about the next Internet-based course in physics that will come in the near future. Using the experience that we have gained now, we are looking forward to improve our methods and organization with regard to distance educational projects. Furthermore the MSU Virtual University, in an cooperation with the Department of Physics and Astronomy, is working on a complete curriculum for an on-line Remote Master's Program in Beam Physics.

References

[Adobe, 1996] http://www.adobe.com/supportservice/devrelations/PDFS/TN/PDFSPEC.PDF

[CAPA, 1997] http://www.pa.msu.edu/educ/CAPA/

[CU-SeeMe, 1997] http://cu-seeme.cornell.edu/

[IBM Alpha Works, 1997] http://www.ics.raleigh.ibm.com/ics/techexp.htm

[Kashy, E., Morrissey, D.J. & Tsai, Y. 1994] *CAPA - System Description and User's Manual*, MSUCL-925. Kashy, E., Morrissey, D.J., Tsai Y. & Wolfe, S.L. (1995). *CAPA - A Versatile Tool for Science Education*, MSUCL-971.

[PictureTel, 1997] http://www.picturetel.com/

[RealAudio, 1997] http://www.prognet.com/

[TCI Software, 1997] http://scinotebook.tcisoft.com/scinotebook/default.htm

[VCS, 1997] http://vcs.es.net/vcs

MASS MEDIA AND IDENTITY

Karin Blair
8, av. Adrien Jeandin
1226 Thonex,
Switzerland
TEL& FAX: (41 22) 349.11.01
E-mail: BLAIR@uni2a.unige.ch
http://www.wfi.fr/blair

Mass media by definition remove the immediate context of any communication such that it can be transmitted by electrical impulses to others at various times and places. This in theory validates our Western idea that all men are equal and therefore communication among them should be so. This assumption of universal human value is distinctly Western, having originated with Las Casas in the 16th century when he wished to stop the exploitation of natives in the Americas by explorers who assumed they were commodities without souls. He counterclaimed that as humans they had souls or at least the right to soul development. Based on this assumption we have four communication models as elaborated by Gordon Pask (Cybernetics and Systems, Vol I&II, ed: Robert Trappl, London: World Scientific, 1994)

One to one:

Individual conversations involve individuals in immediate oral interchange with other individuals; the telephone can extend such communications across spatial distances without inherent limit.

One to many:

The lecture is the traditional means whereby one person addresses many; first the radio and then television take such communications out of their original context and extend them to anonymous audiences many of whom may be distant from one another and unknown to each other. First came radio which left the visual imagination free. Then came television with the power of the image with its claim to veracity. Both have the intimacy of coming into the home but the difference is important . Would Reagan have been so popular on radio and would FDR have been so popular with TV?

Many to one:

The vote or poll exclude feelings and the impact of questions. Many will come up with an answer if asked a question they may never have thought of before. These are ways in which the many can express an opinion to or about a single person; once again media can extent the range of such a model. And the speed of media can influence polls. If there is an 8 o-clock poll followed by newsofits results, then different answers may appear on a later poll.

When mass media are involved the "one" can become an unseen production team. In the absence of a local context of origin, new contexts are created; television programs provide the illusion of immediacy and intimacy, the sense of realism is enhanced. Whereas in the original form of these communication models the context would have been inseparable from the communication, with mass media the context used, however realistic it may look, has been created for specific, although often unspecified, purposes.

Computer networks linked by modem connections permit a new communication model: many can talk with many. This is a mode of communication opened up by technology which can include feelings and other forms of feedback which would have been impossible in the many to one model. At the same time there can be a great deal of anonymity.

These observations fall in the context of the distinctly Western assumption of individual human value, which Las Casas in the 16th century implicitly extended to all humans when he wished to stop the exploitation of natives in the Americas by explorers. The conquistadors assumed they could enslave or otherwise dispose of the natives as if they were commodities without souls. Las Casas counterclaimed that as humans they had souls or at least the right to soul development.

Concomitant with the notion of universal human value which underpins many to many communicaiton is the idea of equality regardless of gender, age or social origin, which led me to the following experiment.

During the EARDHE (European Association for Research and Development in HIgher Education) Conference in 1988 in Utrecht Holland a simulation was organized to explore teleconferencing as a means of interaction. Although computer technology is often thought of in terms of what it can do for us, I wanted to explore what it can let us understand about our everyday operating assumptions affecting communications between individuals. For the first time technology gives us the opportunity to send and receive messages without reference to the parameters of gender, age and social origin, which formerly were necessarily disclosed by a name, a voice and/or a body. The objectives of the simulation were to examine these issues directly by participants reflecting on a hypothetical quota system in which the distribution of gender, age and social origin in any profession would parallel that in the population at large and indirectly by exploring how we assess material received from others known to us only by number.

Communications

What was at that time EARN (European Academic Research Network), linked to BITNET in the USA, was used for the experiment. The groups/individuals which expressed interest in the project were more numerous than those who finally participated. Attempts were made to connect to EARN/BITNET by people in Brussels (Belgium), Geneva (Switzerland), Leeds (England), Nimegen (Holland) and in New Jersey, Washington, Illinois, New York, Colorado, Massachusetts and Pennsylvania in the United States. Those in Geneva, Leeds, New Jersey, Washington, Illinois and Pennsylvania succeeded in getting through. Finding local information on EARN/BITNET was an initial problem everyone did not resolve. Three groups attempted to use regular telephone lines or national packet switching systems but did not succeed in making contact during the conference. There was also the proble of gateway addresses which linked EARN/BITNET to other networks but which would function in only one direction, hence, for example, a message could be received but not answered. The groups which made contact during the conference still encountered the problem of delays in the transmission of messages, with the last paper sent to us on Friday arriving Monday after the conference had ended. Finally those messages which were transmitted during the conference occasionally had crucial words/phrases missing or material present which was meant to have been deleted. Nonetheless the conference did take place and participants there learned enough to justify the effort and more importantly to lay the basis for extending this initial experiment into a longer term project.

Steps in Participation

Participants in different places were asked to prepare an initial position paper taking a stand on a hypothetical quota system. These texts were sent to Utrecht where the biographical information on gender, age and social origin was removed before position papers were redistributed. Participants receiving this material were then asked to formulate a response and in doing so to imagine who they were communicating with. In spite of the various problems encountered, commentaries were made on anonymous and coherent material with some rather interesting results.

Time Frame

The first information on the experiment was distributed in October in order to assess the potential interest in such a project. In February the text, procedures and objectives were mailed out to about forty persons who had expressed interest and in March the following outline was distributed on how to proceed with the presentation of initial positions, the preparation of responses to others and of commentaries on the material. The following outline includes the text, which sets up the hypothetical situation, and the issues for reflection. of your country's government. Your role is to counsel the executive head on the desirability of introducing a quota system. A strict quota system would have the distribution of jobs in any profession reflect the distribution of age, gender and social origin in the population at large. In establishing your position specify 1) which of the parameters should have the highest priority, and 2) whether or not the system should be thought of as universal in application or as targeting certain professions, and if so, which ones?

Interactions and Assessment

Of the various topics proposed the issue of a quota system was used because it raises explicitly problems of gender, age and social origin, which the medium of telecommunication removes from the message. In so far as culture as been conceived of in opposition to nature, the act of thinking of a world structured on a quota system might make cultural arrangement more responsive to the natural factors of age, sex and birth. In addition the skew revealed by comparing the "real" world with a hypothetical one structured by an absolute and universally applied quota system can return us to these real worlds with an enhanced sense of how they are shaped and how one can cope with their skews. Therefore the subject is itself of interest. Of fourteen position papers received, three of which represented fictional characters, seven women and two men were favorable to the establishment of some sort of quota system as a means of redressing imbalances. Two other men were unfavorable on the grounds that stereotyping would be fostered and level of competence diminished. The three fictional characters were created by women and represented two men and one woman each of whom was against any form of quota system.

Of the eight people who made commentaries on texts, three were women and five men, only one of whom continued to express his opinions against a quota system while assessing another's position. When dealing with "real" people, half of the assessments of gender were wrong with one of the two participating men taken to be a woman and two of the four women taken for men. On the other hand for the fictional characters all of the gender assessments were correct. When one moves from gender to social origin one sees how simple an either/or distinction gender offers and how easily subject to stereotyping it can be.

Since social origin could be interpreted in many ways, an incorrect assessment was taken to be any factor which was in direct contrast to the actual situation of the respondent. For example, in one instance someone accustomed to responsibility was taken to be a "back-up" person, in another people from mainland China were taken to be from Western industrialized nations thus reversing easy East-West polarizations. In one case a person from China was thought to be from London or Montreal, which also calls into question easy assumptions about native language perception. Another Chinese was identified with the conservative Christian Party, which reverses notions of group affiliation. Here, too the estimates were wrong half of the time when dealing with real people and correct when directed at fictional characters.

Although all but two of the persons involved in the experiment were middle-aged and therefore provided an easily targeted category, the two younger persons were taken to be older and if "naive" can be said to define an age category, one "older middle-aged" woman was taken to among the young. The fictional Reagan, however, was taken to be middle-aged. Of the three who commented on how it felt to communicate using this technology, one stressed the ease of expressing anger, another the temptation to respond irresponsibly, whereas a third said she felt uncomfortable but responsible.

Although these latter comments are scanty at best, they indicate directions to be explored. In discussion at the conference I met people interested in seeing results of the simulation in the perspective of the impact of Minitel on the general population in France and of that of electronic mail on office relations.

Similarly, in trying to reach some overview of the results of simulation, I find the obvious conclusion to be the need for more work. Since several easy common sense understandings of the world were frequently upset by the few interactions possible during the conference, only two assessments of real people were correct, there is room for future simulations aimed more precisely at specific issues.

Because gender is such an easy dichotomy around which to organize the world, it seems to have been much used and abused. Masculine and feminine as cultural categories imposed on men and woman in the form of stereotypes result in much misplaced concreteness.

Unclear thinking, for example, was a cue to interpret a person as female, used once by a man (A) about a woman, once by a woman about the same man (A) whom she therefore took to be a woman. He, (man A) was taken by another woman to be a typical liberal male with the usual hard-headed chauvinist views that devalued such characteristics as feminine sensitivity. She was surprised, however, to find that he had been so misinterpreted by another woman, who therefore was contradicting the stereotype of superior feminine intuition. While we may need these distinctions for clear thinking, their common sense application which says that males are masculine and females feminine could be called into question.

Future Implications

Two separate issues can therefore be singled out of these interactions. The first could examine on what basis we make these gender distinctions and with what kind of accuracy? Is there a feminine versus a masculine language? way of thinking? A second could focus on how men and women as respondents create images of whom they are speaking to. Are the same stereotypes used in the same way by both sexes, as happened at the conference?

A second issue involving stereotypes was raised by the greater ease with which fictions were identified. Since two of the three fictional characters were mine, I can say that in creating them I was simply remembering specific people, one from twenty years ago, the other from five years ago. Both would have had strong positions on quotas which I tried to imagine without any concern for coherence of position. Consciously I was not dealing with stereotypes and yet,both characters were assessed correctly in all ways by people from very different cultures. Therefore the role of stereotypes in memory has to be raised as well as their role in creating fictions. An additional question might be that of the role of fictions in making the maps by which we navigate in our environment. Without the usual parameters of gender, age and social origin as indicated by the usual cues of name, voice and body, there seemed to be considerable identity confusion, which intersects with the current notion of gender as a chosen performance which can be negotiated and modified.

A third range of issues centers on culture. First of all is there a cultural grid such that individuals from certain ones interpret more or less easily material originating from certain other ones? Do some cultures "screen out" certain others? or reinforce them? If so, on what parameters? under what circumstances? Are there consistent cues which let us apply the main dichotomies we use: east versus west, north versus south? Or like gender distinctions should these be seen as polarities representing possibilities for all people? Related to this is the issue of language: are "native speakers" discernible when the usual cues of name, voice and body are removed? Does this cast light on the role of accent and body language in certain communications, (which ones, under what circumstances?)

Finally, the absence of the possibility of an interactive hook-up among participants raised the question of how the assessments might have been different had the participants been using a conversational mode of interaction instead of dealing with formal positions. To what extent is the taking of a position creating a fiction? Or on the other hand to what extent does the thought process involved in creating consistency blur stereotypical distinctions? Would participants have been easier or harder to identify had they been conversing? What would be the strategies used to get the other to reveal him/herself? Under what circumstances would the need for this information seem most important?

Speculative Extensions

Communications technology like transportation technology is taking us out of familiar territory. Can the generation of alternatives lead us beyond restrictive notions of exclusive values (zero-sum games in which I win because you lose) to scenarios or networks of "mutual-sum" games in which we all share each other's well-being? A personally important question for me is can this model of three intersecting axes function as a gyroscope in this new environment? Although technological innovations may seem threatening, perhaps it is in such alternative worlds that we can find a rich and deep mirror for the people we have been, are and can become in interaction with others in latter years of the twentieth century.

MusicWeb, New Communication and Information Technologies in the Music Classroom

Carola Boehm, Dept. of Music, University of Glasgow, United Kingdom, carola@music.gla.ac.uk

Celia Duffy, Dept. of Music, University of Glasgow, United Kingdom, S.Arnold@music.gla.ac.uk

The European consortium "MusicWeb" came into existance 1995 with the objectives to take an inventory of the problems in computer-aided music education and to produce software solutions for the future years. It provides a growing user community with an library of educational applications for the use in schools, universities or private homes, as well as a digital resource collection of items, and tools out of which educational applications will be able to be build by the users themselves.

To conform to the emerging standards and practices of digital archives and libraries, the Musicweb consortium will cooperate with the "Performing Arts Data Service" (PADS) for storing and archiving music-related items. For the authoring tools, software systems from the hypermedia world are utilised and new pedagogical tools for music-educational purposes are in development. For more information on the project, see http://sun1.rrzn.uni-hannover.de/MusicWeb

Dynamic Web Access for Collaborative Writing

Jennifer J. Burg, Assistant Professor of Computer Science
Wake Forest University, Winston-Salem, NC USA 27109
burg@mthcsc.wfu.edu
Anne Boyle, Associate Professor of English
Wake Forest University, Winston-Salem, NC USA 27109
boyle@wfu.edu
Yinghui Wu, Graduate Student, Dept. of Mathematics and Computer Science
Wake Forest University, Winston-Salem, NC USA 27109
wu@mthcsc.wfu.edu
Yue-Ling Wong, Academic Computing Specialist, Dept. of Chemistry
Wake Forest University, Winston-Salem, NC USA 27109
ylwong@wfu.edu
Ching-Wan Yip, Academic Computing Specialist, Dept. of Physics
Wake Forest University, Winston-Salem, NC USA 27109
cwyip@wfu.edu

For college composition students, the pre-writing and brainstorming processes can be facilitated through dynamic access to the World Wide Web, which offers timely information as well as a channel of communication between writers working on shared themes. Our Writing Tutor program, implemented in Java, uses dynamic Web search to present relevant information to the writer and connects writers through Web-based discussion forums and electronic chat sessions. The writing instructor can offer focus questions to help guide the development of ideas in the discussions. Chat sessions can connect writers at distant locations, making possible an exchange of different cultural perspectives. These enhancements to pre-writing activities are packaged together with the traditional text-editing features of a simple word processing program, creating a single and consistent environment for essay development.

The National Ergonomical Information Network of Ukraine

Alexander Burov
Department of Ergonomics, National Research Institute for Design, Kyiv, Ukraine.
E-mail: burov@ergon.freenet.kiev.ua

The Government of Ukraine has accepted the decree about development of national system of design and ergonomics. Therefore the necessity of development of concept of information maintenance of ergonomics, main conceptual rules of strategy and tactics of developments of national information services of ergonomics in Ukraine has arisen.

Such approach requires to ensure the integration of distributed information resources, storing in the main and regional computer centres and in local computer networks, on enterprises and in institutions with the purpose of maintenance of reference opportunity to information resources of network components as more high, as other components, and international data bases. The interaction between divisions of information-methodical centre and other ergonomical services (including, branches) permits to all services operatively to exchange by all necessary information, supporting the information ergonomical space on all territory of Ukraine. This space make: technique of ergonomical researches, technique of ergonomical examination, methods of prognostigation of requirements in ergonomical services, technology of ergonomical designing (ergo-design technology), banks of ergonomical data, automated workplaces of ergonomist-designer, ergonomist-researcher and practical ergonomist.

The most effective way of decision of such task is the creation of network of Intranet type, which has communication with Internet, and the information circulates within the limits of firm, branch, region and countrie.

Graphical Representation of Students' Laboratory Marks on the World Wide Web

Angela Carbone
Department of Computer Science
Monash University
Australia
Angela.Carbone@cs.monash.edu.au

This poster reports on a JAVA applet called G-SPI which allows students and lecturers to graphically represent student laboratory results and other performance statistics on the World Wide Web.

The traditional system

In the Department of Computer Science students are required to complete a set of programming core units. These units consist of lectures, tutorials and practical classes. Part of the assessment for the core units occurs in the practical classes and results are entered into a database on a weekly basis. Towards the end of the semester administrators generate reports detailing the student laboratory results which are posted on the notice board for the students to review.

Problems associated with the traditional system

1. Students are unable to easily check their results during the semester,
2. Lecturers are unable to quickly detect the overall students' progress,
3. Comparisons between different labs, or individual students to class averages are difficult to make.

Goals of the G-SPI

1. Allow students to check their results regularly throughout the semester and compare their performance with the rest of their class,
2. Provide teachers with mechanisms to detect and monitor student performance quickly and easily,
3. Eliminate the need for posting laboratory marks on the notice boards.

Selecting the Right Person For the Job - An Interactive Tutor Recruitment package

Angela Carbone
Department of Computer Science
Monash University
Australia
Angela.Carbone@cs.monash.edu.au

In 1996, the Department of Computer Science, Monash University, implemented an interactive tutor recruitment package on the World Wide Web. The package was developed to provide:

postgraduate students with information to make an informed decision about front line teaching
useful information to select the right person for the job
structures to pre-plan the semester's teaching activities
an automated timetable scheduling process

This poster will highlight the five sections of the package which include:

1. **Are you eligible to tutor?**
Outlines the personal requirements and qualifications needed by a candidate.

2. **List of subjects available**
Lists the subjects requiring front line teachers in each semester with links to external sources of information detailing the course description, a synopsis of the lecture content, lecture times, consultation hours, etc.

3. **Terms and conditions**
Describes the teaching duties and administrative responsibilities expected of front line teachers and a table detailing remuneration rewards for each type of teaching activity.

4. **The tutor recruitment form**
Uses an interactive form in which each postgraduate enters their personal details, the subject they wish to teach and a ranking of preferred teaching times.

5. **Automatic timetable scheduling process**
After completing the form a report containing a list of teachers with their details is produced for each subject. This report is used to create timetables based on the traditional scheduling methods, such as First-Come-First-Serve Method and on a priority basis.

Open Standard Content Cookies: Utility vs. Privacy

Chad Childers, Ford Motor Company, U.S., chad@ford.com

Iain O'Cain, Intranet Org, Canada, ec@intranet.org

Linda Bangert, Internet Education Group, U.S., lmb@ieg.com

Mike O'Connor, Silicon Graphics, Inc., U.S., mjo@dojo.ford.com

There is an increasing requirement to retrieve more data on individual Web users for delivery of user-sensitive pages and acquiring marketing information. Open Standard Content Cookies allow content negotiation by embedding personal profile information in a request.

How do we deliver tailored content to users while protecting their privacy? The Open Profiling Standard [1] is geared toward gathering marketing demographics, while information on an individual user is encrypted. A central clearinghouse can then provide a demographic report back to content providers, without compromising the user. Giving users granular control of what a site knows about them will head off even more abuse. Profiling oriented towards interests and education, such as Geek Codes [2] can enable site managers to better target their products, while not exposing personal information such as address/ZIP, which is still state of the art for advertisers and online services.

OSCC is an enabling technology for even more personalized websites and a better return on investment. However, electronic privacy advocates have immediate and valid concerns which make implementation challenging. Technologies like DoubleClick have already shown the potential for abuse [3]. OSCC was envisioned as a way of providing profiling information without unduly compromising privacy. We need to continue work on this topic, or market forces like CyberPromo [4] will drive a solution for us as we deliver more tailored content via the Web.

[1] Netscape/Verisign Open Profiling Standard at http://www.firefly.net/OPS/OPSrelease.html
[2] The Geek Code at http://krypton.mankato.msus.edu/~hayden/geek.html
[3] Cookie security risks at http://www.genome.wi.mit.edu/WWW/faqs/wwwsf7.html#Q64
[4] CyberPromo "Spam Iz Gud" at http://www.cyberpromo.com/

A Framework for the Comparison of Computer-Supported Collaborative Learning Applications

Ruth Crawley & Jerome Leary
Collaborative Computing Research Group
Department of Mechanical & Manufacturing Engineering
University of Brighton
United Kingdom.
R.M.Crawley@bton.ac.uk
J.J.Leary@bton.ac.uk

"to my knowledge, we do not have any agreed framework for comparing and contrasting...collaborative learning." [Bannon, 1989]

Today, in 1997, such a framework still does not seem to have emerged. We approached the task of developing a framework of this nature by collating a number of categories by which they might be compared and, through conducting a comparison of thirteen such applications, determined which categories were applicable and which, if any, were missing. In this way we identified five main categories of CSCL application determined by the learning activity supported: tutorial, problem solving, simulation, debate or modelling. Across these categories we found three subsets of feature describing technical, collaborative environment and collaborators' characteristics. Through our investigation a relationship was discovered between the learning activity supported and the pattern of technical, collaborative environment and collaborators' attributes so that CSCL applications supporting, say, modelling activities, displayed a characteristic distribution of attributes. We suggest that such a framework could be used to produce a development tool to determine the specifications for creating CSCL applications to support particular learning activities.

References

[Bannon, 1989] Bannon, L. J. (1989) *Issues in computer supported collaborative learning*. In: C. O'Malley (Ed.) Computer Supported Collaborative Learning. Springer-Verlag 1995, pp. 267-282.

Acknowledgements

This project has been funded by the Joint Information Systems Committee's (JISC) Technology Applications Program (JTAP).

World Wide Web and Data Base Managers: Threads Woven into Fabric

Gregory M. Dick, Information Technology, University of Pittsburgh at Johnstown, USA, gmd@pitt.edu

J. Jeffrey Sernell, Administrative Computing, University of Pittsburgh at Johnstown, USA, jsernell@upj.pitt.edu

Scott Koontz, S. E. Koontz Software Solutions, USA, koontz@vms.cis.pitt.edu

Introduction

The World Wide Web enables ubiquitous communications. Data Base Systems enable the organized storage and extraction of data. These applications can by combined in a synergistic fashion to produce powerful tools which contribute to the solutions of Web communications and data management problems. This paper is a case study of several application of this approach at the University of Pittsburgh at Johnstown (UPJ).

Stale Web Information

Stale information is an embarrassing problem for Web site managers. All too often, the solution involves the regular manual update of HTML files. Frequently, the updated information must also be entered into another related system. Pitt-Johnstown has implemented a Campus Event Scheduler which avoids this situation by applying a Web enabled data base system to the problem. The system is built using:

- Filemaker Pro - a data base manager which stores the campus event data and extracts necessary fields in response to queries.
- WebStar - Web server software for the Macintosh
- Web FM - cgi "middleware" which couples the data base information to the Web.

The Pitt-Johnstown Home Page contains a link titled "What's Happening at UPJ". This link does not lead to a static HTML file which requires regular manual updates. Rather, the Web inquiry is captured by Web FM and converted into a query of Filemaker. The extracted data is "formatted" by Web FM and transferred to the requesting client. The features of this system include:

- Event data is entered into the data base once
- The Event Home Page is always up-to-date because it is constructed, in real-time, from the current data base contents
- The page is self-managing. No manual action is required to keep the page up-to-date.

The result is a very effective page (always contains current information) which is efficiently maintained (automatically).

Remote Access to Applications

The ability to access applications remotely enhances their usefulness. The Information Technology Help desk at Pitt-Johnstown is supported by a Microsoft Access data base. The data base system organizes the management of "trouble tickets". This approach is not new; a long history exists concerning the use of this technique to:

1. Prioritize requests for assistance
2. Track the progress of active tickets
3. Evaluate the performance of the help desk.

However, 3 above required that all help activities be recorded in the data base. This can be problematic. E.G., analysts are occasionally (frequently) "caught in the hall" between tasks and asked to help solve a "little, quick" problem. While the capture of an analyst in this manner may be acceptable, it often disrupts the accurate recording of the help experience. Since it is unlikely that a client database system is available on every computer in the organization, the help experience will be recorded only if the analyst remembers to enter it into the system when she return to her desk later in the day. Busy analysts may forget to do this. If access to the data base was ubiquitous, then the analyst could generate a trouble ticket from the computer of the user being assisted. Ubiquitous access to the trouble ticket data base has been provided at Pitt-Johnstown through the marriage of Web and data base technologies. The system has been implemented using:

- Microsoft Access to organize and access the trouble ticket data
- Windows NT Web Server tools including:
    1. NT 4.0 Workstation - acting as a database server
    2. NT Peer Web Services
    3. NT Internet Database Connector scripting which implements SQL commands against the Access Data Base.

The result is a trouble ticket data base which is accessible via a password protected web form. Analysts may write trouble tickets, or query the data base from virtually any computer on campus.

Gathering Data

Data base systems are excellent tools to analyze and summarize survey data. A problem can occur in the data acquisition phase. A typical process might be:

- Survey participants complete a form
- Form data is entered into the data base
- Analysis is performed and summary results determined.

The first two steps, gather and enter the data, are time consuming and amount to double data entry (i.e., both the participant and the data entry clerk handle the same data). The use of machine readable forms decreases the effort involved, but a significant amount of manual effort is still involved (gather forms, execute data entry application, feed forms to the reader, etc.). A better solution would allow the survey participants to directly interact with the data base. Such a system has been implemented at Pitt-Johnstown as part of the colleges Freshman Network. During Freshman Orientation, new students are formally introduced to a wide variety of offices and services at the college. Included in this program is surveying the entering students about a variety of topics so that their needs might be better served. This survey procedure supports direct data input, by the students, to the data base. The implementation involves the use of a Web form which is coupled to the underlying data base manager by standard Windows NT tools. No intermediate handling of the data is required.

Conclusions

These experiences lead us to conclude that commercially available Web tools and data base systems can be woven together into data management systems which are both:

- Powerful - because of the underlying data base engine
- Ubiquitous - because Web browsers are available virtually everywhere.

This combination of tools is not new; Web search services are built using a similar combination of tools. Their use at Pitt-Johnstown has led to highly effective and cost-effective solutions to problems involving both the need to provide up to date web data and the need to gather information from a dispersed set of locations.

Vocal Point: A Collaborative, Student Run On-Line Newspaper

Scott Dixon
Teacher, Centennial Middle School
Boulder, Colorado, USA
sdixon@bvsd.k12.co.us

Chrissy Anderson
Student, Centennial Middle School
Boulder, Colorado, USA
andersoc@bvsd.k12.co.us

Many educators see access to the World Wide Web as a resource for their students to do research and gather information to be used in the classroom. The power of the Web also allows students to become providers of information - sharing research findings, linking their analysis to valuable "live" references and presenting their thoughtful opinions on real world issues using this powerful presentation media.

Vocal Point *(http://bvsd.k12.co.us/schools/cent/Newspaper/Newspaper.html)* is an award-winning, on-line, collaborative, electronic newspaper created and managed entirely by students. It is produced by students from Centennial Middle School *(http://bvsd.k12.co.us/schools/cent/CentennialHome.html)* in Boulder, Colorado, along with students of all ages from around the world. Vocal Point was the first of its kind in the world and continues to showcase leading edge technology that highlights student research and viewpoints.

Utilizing the WWW for Industrial Training

Eman El-Sheikh, Fran Bakowska, Chris Penney, Rong Liu, and Jon Sticklen

Intelligent Systems Lab
Computer Science Department
Michigan State University
East Lansing, MI, USA
email: elsheikh@cps.msu.edu or sticklen@cps.msu.edu

A web tutorial is a cost-effective method for the education and training of people in the manufacturing industry as well as in other environments. An Internet-based tutorial can be used by many users simultaneously at different locations. In addition, it allows each student to browse and progress through the material in his/her own order preference and pace. We describe a web tutorial for the domain of polymer composite molding, a field of importance to the chemical engineering, material science, and mechanics communities. The intended users include industry professionals, university students, and others interested in learning about the domain of polymer composite molding. Several navigation techniques are provided to allow users to access the information in the most suitable manner. Users have the option of navigating directly to the specific information required or reading through the tutorial sequentially in a step-by-step manner. To make the learning process more effective, the tutorial utilizes various media forms to present the information to the user. The latest version of the tutorial is available at http://isl.cps.msu.edu/trp/.

A Practical Approach and Infrastructure for Large Scale Web Applications and Services

Pierre France and Michele Kearney
Systems & Computer Technology Corporation
4 Country View Road
Malvern, PA 19355
(pfrance, mkearney)@sctcorp.com

Abstract: This paper describes SCT Aspire, an online learning infrastructure for large-scale web applications. The infrastructure is for corporations and training companies that employ thousands of learners. The infrastructure enables you to author, administer, and manage the learning experience on the web. The paper starts with a short discussion of the problem domain, web technology, training services, and an approach to address the issues. Next, the paper demonstrates a series of maps that outline the conceptual design, logical site architecture, and physical architecture used to construct the overall infrastructure. Finally, a discussion of front-end user interface design is discussed as well as technology issues relating to implementation.

Identifying the Problem

The World Wide Web has a number of limitations that create obstacles for online learning. First, there is a lack of standard protocols for constructing and designing a web architecture that allows users to be instantly familiar with the interface functions. There is no way of knowing in advance where a particular link or navigation aid will take the user. Current web sites are designed to navigate many ways via blind links, search-engine requests, and drill downs to investigate information. There is no single right way to do this. Next, users accessing courses and their learning objects are slowed by the rate of the bandwidth to navigate and interact with the material. To make static HTML pages more interactive requires expertise in CGI, Java, and ActiveX as well as the knowledge of current plug-in technology. In addition, new push technology [Kelly & Wolf 1997] is developing quickly hence, making it difficult for the average web developer to stay current. Lastly, individuals that have knowledge and skills in maintaining networked web servers, object-oriented databases, and electronic commerce are difficult to find.

With this in mind, very few software products in online learning address the two key problems--administration and site management. How do learners register for courses and pay externally? How do you manage 20,000 learners and hundreds of online courses and their learning objects (test questions, objectives, e-mail messages, schedules, annotated notes, assignments)? And how can you guarantee Internet performance to learners and keep current with hardware and software upgrades?

If corporations developed an online learning system much of the efforts would focus on building a large-scale system that authors, administers, and manages learning. Next, they would turn their focus on having to learn how to convert traditional courses to online, maintain a web site, generate various tracking reports, and set up the organization's billing process. It would also require hiring uniquely skilled people in the area of information technology and other disciplines. This is beyond the expertise, time, and budget for most organizations.

One Solution Approach

The challenge: Deliver a single source solution for corporations and training companies with online learning needs. A solution that helps corporations reach more learners to improve job performance. A solution that also helps training companies create new markets and protect existing ones. In addition the solution must reduce training and education costs, allow organizations to customize courses, and provide comprehensive administration.

The solution: SCT Aspire, an online learning solution for large-scale applications to deliver the learning experience via the web.

Goal Definition

The goals of SCT Aspire are: (1) provide corporations with one solution to author, administer, and manage the learning experience via the web; (2) provide options for web site management to guarantee Internet performance and keep current with hardware and software upgrades; (3) provide all functionality to the desktop requiring no more client-side intelligence than an Internet browser; (4) allow users to pay online for their experiences via a credit card; (5) allow multiple entrances to the site for minimum navigation and maximum flexibility; (6) provide a motivational and interactive experience utilizing the most modern organization and presentation methodologies; and (7) provide full support services for business planning, content creation, and custom applications.

Conceptual Design

Given the goals, there are seven main areas of functionality included in the infrastructure [Fig. 1]. These are: (1) Content Conversion Services - services for converting passive learning content (text, audio, video and images) into effective interactive and motivational content for deployment; (2) Web Site Hosting Capabilities - develop and maintain web sites for users' learning content, business information, products, and services; (3) Billing via Credit Card - potential users can review the course catalog and pay for courses online with a valid credit card; (4) Group Communication - provide real time collaborative tools for users such as e-mail, chat forums, bulletin board technology, document conferencing, and personal scheduling; (5) Database of Users - collect user information to provide detailed tracking reports on learner activity, courses, and billing records; (6) Team Training Companies with Content Providers - partner with training companies and content providers to develop unique online services and products for users; and (7) Single Solution - provide users with a single source for all of their web-based education and training needs.

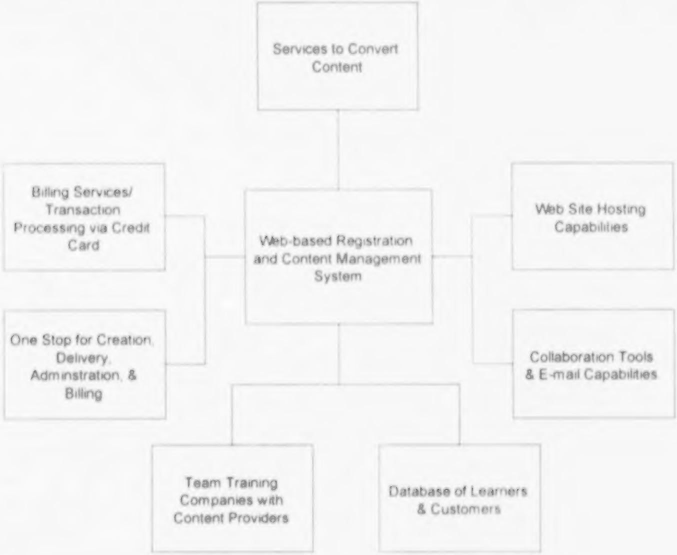

Services to Convert Content
Billing Services/ Transaction Processing via Credit Card
Web-based Registration and Content Management System
Web Site Hosting Capabilities
One Stop for Creation, Delivery, Administration, & Billing
Collaboration Tools & E-mail Capabilities
Team Training Companies with Content Providers
Database of Learners & Customers

Figure 1: Conceptual Design

Logical Site Design

While the site's overall goals, messages, and content structure are part of the conceptual design, the logical design is represented by a site architecture diagram and technical specification.

The site architecture [DiNucci et al. 1997] describes the organization of the content, relationships of the pages, and links of the site. The site architecture diagram for the project is shown below [Fig. 2].

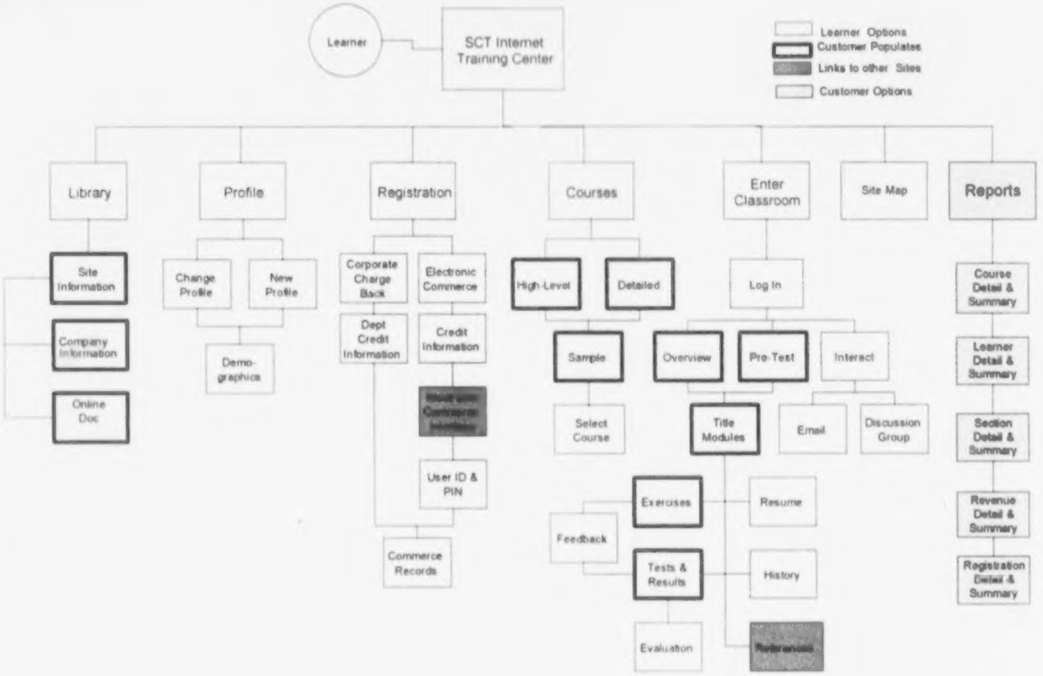

Learner
SCT Internet Training Center
Learner Options
Customer Populates
Links to other Sites
Customer Options
Library
Profile
Registration
Courses
Enter Classroom
Site Map
Reports
Site Information
Company Information
Online Doc
Change Profile
New Profile
Demo-graphics
Corporate Charge Back
Electronic Commerce
Dept Credit Information
Credit Information
User ID & PIN
Commerce Records
High-Level
Detailed
Sample
Select Course
Log In
Overview
Pre-Test
Interact
Title Modules
Email
Discussion Group
Exercises
Resume
Feedback
Tests & Results
History
Evaluation
References
Course Detail & Summary
Learner Detail & Summary
Section Detail & Summary
Revenue Detail & Summary
Registration Detail & Summary

Figure 2: Logical Site Architecture

The diagram is the first step in the logical design of the site and accompanying applications. From the site architecture, possible presentations that would be most effective to meet the site's goals are visible.

Also, the roles of the various site users need to be defined. For this work, the two principle roles are the learner and the customer. The learner becomes the training center member and experiences the courses. The customer is the company responsible for providing the content of the training material. Therefore, the learner is the site's primary user and design elements and considerations will be made with the learner's interests in mind.

Each of the goals for the site needs to be represented on the diagram. For example, one of the goals for the project is to allow a learner to become a member of the site or register for courses and pay the associated fees via credit card. The Profile and Registration paths clearly show the progression of the pages needed to provide demographic information, register, and pay for the courses to receive a user name and PIN.

Another goal is to allow entry to the site via multiple paths. With a user name and PIN, a registered learner can enter the classroom directly to experience the course without navigating through unnecessary pages. All functionality for both the learner and customer is available with no more on the desktop than a browser.

BEST COPY AVAILABLE

The design is thorough and provides a single solution for creation, delivery, administration, and billing for the customer as well as the learner.

Physical Site Design

To turn logical design into reality, the selection tools and techniques that deliver the design creatively and effectively to the web is critical. The key physical design issues addressed [Fig. 3] are: (1) create a flexible and elegant solution that allows for integration with select third party software over which SCT can influence but not control look and feel; (2) develop an overall approach for a seamless presentation to the end user that will integrate with more than one other pre-developed software package; and (3) define a simple design that anticipates change so that one can update information, add more functionality over time, switch or add integration partners, or customize features when appropriate without total redesign.

Figure 3: Physical Site Architecture

The choice of tools (for example, HTML and Java) will allow the application to run under Internet browsers such as the latest versions of Netscape Navigator and Microsoft Internet Explorer. The use of database-generated pages means that changes can be made automatically and immediately reflected on the web site for flexible change management.

Front-end User Interface Design

Several front-end user interfaces were considered for the SCT Aspire project. The front-end user interfaces selected are a mixture of interfaces that rely on interactivity, motivation, and the learner's experiences. The interfaces include image maps and real-world metaphors. The key to defining the interfaces is to design the solution with the learner's beliefs, wants, needs, experiences, and expectations [Mandel 1997].

An image map [DiNucci et al. 1997] interface was chosen because it presents the user with a graphic interface and a group of links. The map can represent iconic navigational elements throughout a site or a full-page graphic interface. This interface allows the user to select where they would like to go and the order that they wish to navigate. The advantage of this interface is it is much more interactive and users do not need to navigate to areas in a sequence. The disadvantage is that the user may not "enter" all of the areas that they need to completely use the site.

A real-world metaphor interface was also selected for this project. It allows users to transfer knowledge about how things should look and work [Mandel 1997]. If the interface is designed properly the user shouldn't have to learn anything new because the user is familiar with the environment. The environment could be an office, telephone, building, or a classroom. For example, if you are developing a virtual classroom and students know how to use a classroom, they have experience and certain expectations of how a classroom should work. The advantage of this interface is that it is very familiar to the user and it motivates the user to explore.

Implementation

The current architecture work includes design and development of an object repository for the storage, assembly, and delivery of courses. The level of granularity at which the content is stored depends greatly on how the content was created. For example, content that is created from a large paper document scanned into a single TIFF file will not allow the addition of audio or video to the file, the creation of hyperlinks for more interactive navigation, or the storage of content as anything other than a single and large object. At the other end of the spectrum, content that is created with the idea of presenting the user with 20 minute segments of text interspersed with 60 seconds of audio and 20 seconds of video will be created and stored as many separate files. The first example, while providing the least amount of flexibility for the end user, provides little challenge for assembly into the final presentation. However, as the level of granularity decreases and the specificity increases (i.e., content objects become smaller and exponentially more numerous), the final presentation process becomes one of selection and assembly of large numbers of separate files (content objects) in the right order. Object management becomes a critical and nontrivial task as the product begins to support many customers and courses.

SCT provides all content creators with standards for the development of their work that will yield maximum value from the system. Customers will, however, be able to submit content that does not quite meet the standard and the application will provide a limited subset of all of the possible features. This flexibility allows customers to utilize existing content while striving to create new courses using more modern techniques over time.

In Summary

The design of the project enables online registration and assessment, course management and creation, administration and billing, group collaboration, and web site hosting services. This supports the overall goal of being the single solution for all facets of online learning for large applications of learners.

References

[Bass & Dewan 1993] Bass, L., & Dewan, P. (1993). *User Interface Software*. John Wiley & Sons, Inc.

[Collins 1996] Collins, D. (1996). *Designing Object-Oriented User Interfaces*. Redwood City: Benjamin/Cummings Publishing.

[DiNucci et al. 1997] DiNucci, D., Giudice, M., & Stiles, L. (1997). *Elements of Web Design*. Berkeley: Peachpit Press.

[Galitz 1997] Galitz, W. (1997) *The Essential Guide to User Interface Design: An Introduction to GUI Design Principles and Techniques*. John Wiley & Sons, Inc.

[Kelly & Wolf 1997] Kelly, K., & Wolf, G. (1997). Kill Your Browser. *Wired Magazine*, March 1997, 12.

[Kristof & Satran 1997] Kristof, R., & Satran, A. (1997). *Interactive by Design, Creating & Communicating with New Media*. Mountain View, CA: Adobe Press.

[Mandel 1997] Mandel, T. (1997). *The Elements of User Interface Design*. John Wiley & Sons, Inc.

[Shaw & Garlan 1996] Shaw, M., & Garlan, D. (1996). *Software Architecture*. Upper Saddle River: Prentice-Hall.

Acknowledgments

Our visionary - Phillip Clark

Our team – Jamie Battin, Jim Carter, Jane Coyle, Betsy DiCostanza, Rebecca Kolvereid, Nonna Reznik, Dan Sterling, Jennifer Sternbergh, Margaret Ternowchek, and Mark Triest

Our partners – Asymetrix Corporation, MCI, Oracle Corporation

Election Project Experience

Ari Frazão, Fabiola Greco, Lúcia Melo, Teresa Moura
Brazilian Research Netword (RNP)

In a "first" for Latin America, RNP - Brazilian Research Network and TSE (Supreme Electoral Court), in conjunction with the Brazilian Embassy in Washington, worked on a Election Project to make Brazilian municipal election results available on the Internet.

Starting with the close of voting, with updates every 30 minutes in the first round and every 15 minutes in the runoff round, the central computer at TSE forwarded the most current returns to RNP, in its two points with better connectivity, located at São Paulo and Brasília. These two points, in turn, spreads the results automatically across RNP Internet backbone to 7 other points of presence geographically distributed, and also to the Brazilian Embassy in Washington.

Internet users, in Brazil and abroad, were thus able to follow the election returns from web sites distributed across the country. The user was guided to select the fastest option for election return retrieval (which is dependent upon where the user is located).

Designing Web-Based Instruction for High School Courses

Jed Friedrichsen, Instruction Design Specialist
Division of Continuing Studies, Distance Education
University of Nebraska-Lincoln USA
jof@unlinfo.unl.edu

Marilyn Altman, Instruction Design Specialist
Division of Continuing Studies, Distance Education
University of Nebraska-Lincoln USA
maltman@unlinfo.unl.edu

Cynthia Blodgett-McDeavitt, Instruction Design Specialist
Division of Continuing Studies, Distance Education
University of Nebraska-Lincoln USA
blodgett@unlinfo.unl.edu

Distance education is one solution for educational commmunities to reach remote learners, home-schoolers, or otherwise extend traditional curricula. CLASS (Communications, Learning, and Assessment in a Student-centered System), a Star Schools initiative of the University of Nebraska-Lincoln's Department of Distance Education, is developing and testing a complete, accredited distance learning high school curriculum delivered via the Internet.

Targeted toward at-risk learners, CLASS courses are grounded in collaborative and contextual design to engage students in meaningful learning experiences. In a learning environment in which the student may never see the teacher or other students, collaborative experiences include building online learning communities and group-oriented activities.

The instructional design process used by considers diverse learning styles, that inexperience with technology may itself serve as a barrier to learning, best use of the non-linear nature of the Web, and how to balance text with multimedia elements to maximize learning for learners with undeveloped literacy skills.

Text Generation in Business Object Frameworks

M.O. Froehlich and R.P. van de Riet
Department of Computer Science
Vrije Universiteit
Amsterdam, The Netherlands
{frohlich,vdriet}@cs.vu.nl

Emerging Standards for distributed objects like CORBA and DCOM allow the construction of standardized business objects. Currently there are some proposals for sets of such objects (for example from SAP or from CORBA) for different domains. A set of business objects together with a number of rules, defining how they interact is called a business object framework. This kind of libraries is a good source (knowledge base) for automatic text generation. Text generation is a technology that allows to construct texts for a certain domain with different content and style on the fly. The standard object definitions can be augmented with semantic information using a lexicon. This additional information is necessary to use generic tools for text generation developed in computational linguistics. A description how arbitrary conceptual structures are useable can be found in the proceedings of NLDB 97 [1].

References

[1] M.O. Froehlich and R.P. van de Riet (1997). Conceptual Models as Knowledge Resources for Text Generation. In Proceedings of the Third International Workshop on Applications of Natural Language to Information Systems (NLDB 97), Vancouver, Canada, 1997. 1-10.

Pornography is Not The Problem: Student use of the Internet as an Information Source

Katherine J. Fugitt
School of Education, University of Washington, USA
kfugitt@u.washington.edu

The World Wide Web holds great potential for education because of its multi-media and ease of use aspects, though many people worry about keeping kids safe from Internet pornography and predators. Much less concern has been given to the possibly more important (and certainly more prevalent) issue of Internet advertising, marketing, and propaganda and their effects on students. This pilot study begins to identify some important questions about students using Internet resources. Preliminary findings indicate that: 1) students may have difficulty distinguishing advertising and marketing from more "pure" information; and 2) students don't keep in mind that *anyone* can publish *anything* online and don't think about what that means for information reliability. Students aren't being taught to think critically about advertising or the sources of the information they are getting. Student training in critical thinking and media literacy should be further explored as possible ways to help alleviate the problem.

Empirical Analysis of the Use of Electronic Bulletin Boards Supplementing Face to Face Teaching

Jayne Gackenbach
Department of Communications, Athabasca University, Canada,
jgackenb@gpu.srv.ualberta.ca

An evaluation of the use of electronic bulletin boards as part of three traditional face to face classroom offerings for upper level university undergraduate courses in psychology was undertaken. These Augustana University College classes were taught by the author with each using an electronic bulletin board as part of the course requirements. The students were required to post onto the course bulletin board at least 10 out of 14 weeks during the semester. In addition to information gleaned from the students posts other information which sheds light on the value of bulletin boards in traditional classrooms include the sex of the student, major, year in college, and grades in the course. At the end of the semester the students were asked to fill out a questionnaire regarding their use of the bulletin board in class.

Solution Prototype for the Adaptation of an Information System into an Intranet/Internet Environment under Windows95

A. García-Crespo, P. Domingo, F. Paniagua, E. Jarab, B. Ruiz
Computer Science Department, Universidad Carlos III de Madrid
Butarque 15, 28911 Leganes (Madrid)
Tel: 34-1-6249417, Fax: 34-1-6249430
E-mail: agarcia@rioja.uc3m.es

Solution Prototype Description

The aim of this Project is to convert a traditional Information System, by means of a reengineering process, into a functionally equivalent one, but adapted to an Intranet/Internet environement, including execution capabilities trough conventional Internet navigators.

The aim is to show the feasibility to create applications, in compliance with customer requiremennts, following a prototyping strategy. The results will be easily applied to real problems in current information systems.

Objectives achievement must be obtained under an Internet strategy. Results must be observed from any WWW visor site. Internet technologies offer today a real opportunity to build more efficient information systems, as well as competitive advantages in corporate applications.

The main phases of the Project are:

1. Study and analisys of present information system.
2. Definition of new user interfaces.

Figures 1 and 2 shows the same query using previous and current prototype system.

The current prototype includes a history file referring to some relevant parameters, under the same query window. The original system required new queries to access the information, by means of pushing buttons. Data retrieval is the same, but accesibility is better using the prototype.

Observing Figure 1 and Figure 2, the improvement in the query interface is obvious.Figure 1 shows the amount of options and navigation posibilities. Options are not self-explanatory, so we can asume that users will require a large learning proccess.

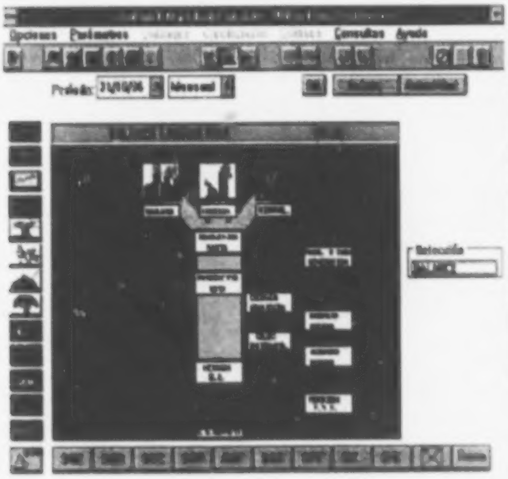


Figure1.
Query: :previous "Balance Energético".

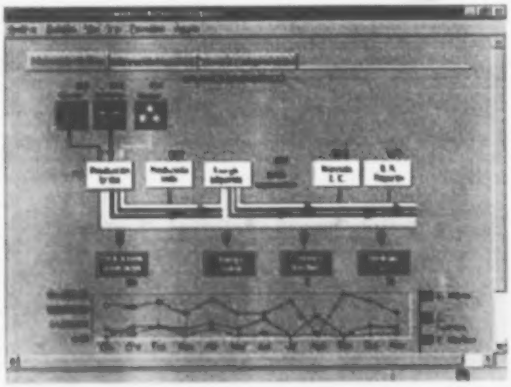


Figure 2.
Query: present "Balance Energético".

BEST COPY AVAILABLE

A New Twist to an Old Idea - Telementoring Using the Web

Melanie Goldman
National School Network
BBN Corporation
70 Fawcett Street
Cambridge, MA 02138
mgoldman@bbn.com

"Telementoring is as an activity where, if many people contribute just a small amount of effort, it can make a big difference in the education of a group of students. " Member, National School Network

BBN's Mentor Center™ is a web-based tool that harnesses the power of the Internet to foster telementoring relationships and thereby vastly expand the number of volunteers that can particpate in and enhance the educational environment. Working with a 6th grade teacher, we initially designed Mentor Center™ as a way to have community members serve as mentors in an ongoing, constructive relationship with students to help them with their writing. Mentor Center™ has evolved to where any type of work available through the Web, text, graphics, sound, can be shared through this tool.

Acknowledgements
Thanks to Henry Olds, Allysen Palmer, and Howie Rafal for their creative talents in the development of Mentor Center™. This project was funded by the National Science Foundation, Grant Number RED-9454769

A Class Web Page Comes to Life

Patricia Gray, Ph.D.
Assistant Professor of Music
Rhodes College
Memphis, TN
USA
gray@rhodes.edu
http://gray.music.rhodes.edu/musichtmls/Music120.html

In 1995, Rhodes' "Music in Eastern Europe" class began a Web page. Among its first projects was a set of papers on Czech music. One dealt with Pavel Haas, a Czech composer who was one of many artists and musicians held at the Terezin concentration camp near Prague. Through a remarkable series of events, the publication of this paper led to students' being able to interview a survivor of the camp. Members of subsequent classes in 1996 and 1997 added further reports, pictures, and sound clips dealing with the Terezin story. In February, 1997, this work resulted in Rhodes' sponsoring a lecture-recital of art songs written on poems by children prisoners at Terezin. The composer, Jeannie Brindley-Barnett, knew of the class project because she encountered it on the Web. This experience points up the unique possibilities the Web presents for making classwork connect to the outside world.

Using Faculty Focus Groups to Conceptualize a Case Seminar Facility for Distance Education Courses at The Medical College of Georgia

Kathleen M. Hannafin, Ph.D., Associate Professor and Director
Office of Educational Design and Development
Medical College of Georgia
Augusta, Georgia
khannafi@mail.mcg.edu

Shary L. Karlin, Ed.D., Coordinator of Distance Education
Medical College of Georgia
Augusta, Georgia
skarlin@mail.mcg.edu

A dilemma often encountered by distance educators is they are expected to "teach" in generic distance education facilities which do not typically support pedagogical strategies such as case or seminar teaching. The design of classroom facilities and hardware and software infrastructures are often determined using a top-down approach, including administrators and technicians, but neglecting to elicit input from end users. Many resulting campus facilities are designed as instructor-centered, with video-intensive technologies (e.g., 2-way ITV). Few facilities on campus currently provide additional software infrastructures for computer or conferencing capabilities to extend pedagogical activities to include web-based case analysis, real-time collaboration, demonstrations, simulations or data retrieval activities.

In an effort to enhance our distance education capabilities to include a wider range of pedagogical activities, a more robust infrastructure and architectural features which support case and seminar teaching, MCG undertook the process of creating faculty focus groups to elicit faculty input regarding the design of distance education teaching facilities.

Integrating On-Line And Face-To-Face Work In Professional And Learning Environments

Don Harben
Toronto Board of Education
Curriculum Department
70 D'Arcy Street
Toronto, ON
Canada
M5T 1K1
Don_Harben@tednet.oise.utoronto.ca

John Myers
OISE/UT
272 Bloor Street West
Toronto, ON
Canada
John_Myers@tednet.oise.utoronto.ca

We are awash with work to be done and with information to do it with. At the same time there seems to be less time and fewer resources available for the tasks. Information Technology can be an effective means to work smarter, provided we are aware of the possibilities and limits especially in terms of using proven group work strategies adapted to use with information technology tools which address constraints of time and place.

In our work with others over the past few years using on and off-line IT applied to common tasks, we are beginning to see some of these possibilities and limits in the complementary use of various Information Tools and strategies.

In this poster session we will share some of our work in progress in the form of a matrix allowing for comparison of a variety of on and off-line Information Technology tools and strategies using group size, task function and work group setting as categories.

We invite feedback for further investigation.

Educators in Instructional Technology

Jeffery L. Hart
Graduate Student
Department of Instructional Technology
Utah State University
USA
jhart@sisna.com

We need to cement the role of the average teacher in the field of instructional technology (IT). The definition of the field is ever changing. The changes that have taken place in the thinking of the field over the last 35 years have helped to mold what educators are today. We need to look at the role that educators play in the field and to look at how they have reacted to the changing nature of technology in education. Education is by nature a slow-moving and slow-changing field. We need to infuse it with the vigor of change associated with today's technological improvements. Instructional technologists can and should be the leaders in education. We need a uniform understanding of IT and a union of its efforts in improving education. Technology needs to improve constantly or it will not survive. Educators must do the same.

A Methodology for Determining Website Navigational Efficiency

Jeffrey B. Hecht
Illinois State University

Perry L. Schoon
Florida Atlantic University

Many factors influence how efficiently a World Wide Web user can navigate a given web site. A systematic methodology, with operational software, was created to allow the study of a user's navigational trail through a web site's pages. The procedure records: the user's most straight-line path (fewest number of steps) from the "Home" page to a given "Target" page, how often the user returned to the site's "Home" page, how many times each page was reloaded, the number of retraced steps, and the number of times each page in the site was accessed (and for how long). One process records the raw information, while another analyzes the recorded data. A computed Navigational Action Efficiency (NAE) index describes user efficiency in navigating the web site; other statistics indicate potentially confusing pages and links. Such information can help web page designers create sites that are easily navigated in addition to aesthetically pleasing.

The Rhythm of the Web: Patterns of "Multiple N's of One"

Andrew Henry, Department of Counseling, Special Education, and Educational Psychology, Michigan State University, USA, henryand@pilot.msu.edu

Valerie L. Worthington, Department of Counseling, Special Education, and Educational Psychology, Michigan State University, USA, worthi14@pilot.msu.edu

The ability of people to communicate is central to social and individual development. This communication assumes many forms (oral, print, and, increasingly, hypertextual) and its developmental impact,, according to a construct known as the Vygotsky Space (Harre, 199?), begins when social knowledge is appropriated and transformed. The next phase is the publication" of beliefs about the knowledge. This publication enters the sphere in the fourth phase, providing the opportunity for others to conventionalize her knowledge. Hypertexts, however, require a reconceptualization. They compel the reader to construct knowledge based on an individual, temporal experience. Here, however, there is no external validation of the concepts that the reader develops as she moves through a hypertext. The publication of transformed social knowledge becomes the reader/author's experience--the path, the links, and knowledge--constructed by the "author" and the "reader". Since this "publication" is private and unique, the whole notion of conventionalization in the public sphere must be reconsidered. Following on the clinical psychology construct of the "n of one," our study develops a sense of the individual experience and the potential for recognizing group or subgroup patterns. Tracking pattern" data, largely unavailable before the advent of the World Wide Web, combined with interview data is an enlightening means by which to begin to understand the motivations, goals, prior experiences of users, to "reconstruct" the meaning they create as they engage with hypertexts, and to come to terms with a new conceptualization of the Vygotsky space.

Acknowledgements

Thanks to Yong Zhao for his assistance in helping us develop these ideas.

Using Internet Tools to Create Cross-Disciplinary, Collaborative Virtual Learning Environments

Ms. Peggy Hines
Office of Distance Learning, Morehead State University, USA. p.hines@morehead-st.edu

Dr. Phyllis Oakes
Elementary, Reading and Special Education, Morehead State University, USA. p.oakes@morehead-st.edu

Mr. Calvin Lindell
Communications, Morehead State University, USA. c.lindel@morehead-st.edu

Ms. Donna Corley
Nursing and Allied Health Sciences, Morehead State University, USA. d.corley@morehead-st.edu

This Poster/demonstration will describe the application of web based tools to accomplish a collaborative project between graduate and undergraduate students across 3 disciplines. Students in each course had a specific role: graduate students identified health-related problems they encounter in the public schools; nursing students researched and reported on the nature of the problem; communication students developed viable action plans. Problems and student outcomes will be identified and discussed. The project's goal was to explore commonly found health issues existing among children and develop strategies to reduce the problems. These tasks were accomplished using technological tools such as a faculty web page, newsgroups, and Nicenet.

Collaborative diagnosing and distance-learning materials for medical professionals

Takahide HOSHIDE, Yasuhisa KATO, Yoshimi FUKUHARA, Makoto AKAISHI, David W. Piraino, James D Thomas, Mitsuru Yamada

NTT Information and Communication Systems Laboratories
School of Medicine at Keio University
The Cleveland Clinic Foundation
KDD

We had two international experiments about collaborative diagnosing and learning materials for medical professionals. These experiments were jointly conducted by School of Medicine at Keio University in Japan and the Cleveland Clinic Foundation in the U.S.

We used 10 Mbps ATM network between Japan and the U.S. At each site we used one Windows95 PC with a MPEG-2 decoder board.

(1) Tele-conferencing for Cardiology: The doctors in both sites discussed using TV conference system and viewing diagnosis echo video images in MPEG-2 format. The capability of MPEG-2 video for diagnosing patients appeared.

(2) Radiological learning materials on the WWW: Both sites had patients' database on the Web, such as History, Findings, and Category, and X-ray photographs in JPEG format. The doctors accessed the databases in Japan and US by using Web browser. It was acceptable to browse the remote server.

UsingWeb Sites in University Courses as Bulletin Boards and for Enrichment

M.Eleanor Irwin, Division of Humanities, University ofToronto at Scarborough, Scarborough, Canada
irwin@scar.utoronto.ca

I developed two course Web sites which were in use throughout the 1996-7academic session, one for CLA A02Y: Greek and Roman Mythology which hadmore than 100 students, the other for CLA B52S: Women in the Greek andRoman world, with 11 students.
URLs
<http://citd.scar.utoronto.ca/CLAA02/CLAA02.html>
<http://citd.scar.utoronto.ca/CLAB52/CLAB52S.html>.

For CLA A02Y, I used the Web site primarily as a bulletin board and postedmy course outlines, assignments, announcements and occasional notes onlecture material and videos as well as slides with instructor's notes.

For CLA B52S, I encouraged student participation by having a discussiongroup set up to which students were required to contribute and havingstudents prepare a simple document (a brief biography of a woman in the Classical world written in html) which was postedon the Web site.

Acknowledgements

The development of the CLA A02Y site was supported bya grant from the Provost's Instructional Technology Courseware DevelopmentFund, University of Toronto.

I want to thank William Barek, Centre for Instructional TechnologyDevelopment, University of Toronto at Scarborough, who designed and managedboth pages.

A path over the Internet to a student-centered on-line classroom: The SUNY Learning Network -- the design, development of nineteen SLN Web-based courses

Mingming Jiang, State University of New York Learning Network, Center for Learning and Technology,
SUNY, Saratoga Springs, Email: mjiang@sln.esc.edu, USA

The State University of New York Learning Network (SLN) is helping SUNY faculty convert traditional courses into Web-based courses, a project funded by the Sloan Foundation and SUNY Office of Educational Technology. In fall 1996, nineteen courses were developed in Lotus Notes and delivered over the World Wide Web through Lotus Notes Domino Web server which automatically translates Notes constructs into HTML for display on the Web site.

This poster session will present in detail the design and development of the nineteen courses, the course interface as well as the ongoing evaluation of the courses. It will present its rationale for design of the dynamic course structure which places important emphasis on interactive communication and higher level learning activities. The purpose of the session is to share our web-based course design and development experiences with, and to invite comments and suggestions from, educators and scholars in this field from all over the world.

Windows to the Universe: An Internet-Based Educational Resource for the General Public (http://www.windows.umich.edu)

Dr. Roberta M. Johnson
Space Physics Research Laboratory, The University of Michigan, USA, rmjohnsn@umich.edu

Abstract: *Windows to the Universe* is an award-winning, NASA-funded World Wide Web site developed at the University of Michigan by a team of scientists, educators, programmers, artists, and museum and library specialists representing several institutions. *Windows* provides an interdisciplinary introduction to Earth and Space Sciences with content ranging from sciences' historical origins to the latest news and research findings. In addition to a wealth of data and factual information, the site emphasizes the artistic and historical connections between science and the human experience as portrayed in mythology, art, music, film, literature, history and philosophy. Content is presented in a multilevel format geared to meet the differing needs of elementary, middle, and high school students. The graphics-intensive layout, easy-to-use navigation tools, and carefully selected content ensure that *Windows to the Universe* is a useful and engaging educational resource for the K-12 classroom, library, or science museum.

Use of Browser-based Technology in Undergraduate Medical Education Curriculum

Laleh S. Khonsari

Medicine in United States is in crisis. Today the challenge to the medical education system is to improve the integration of academic scholarship with an educational process suitable for preparing medical students for the contemporary, service-oriented, dynamic, and demanding practice environment. Integrating medical informatics into the full spectrum of medical education is a vital step toward implementing a new instructional model, a step required for the understanding and teaching of modern medicine. Medical Informatics is the use of technology to provide quality care / education in a most efficient, cost-effective way. It provides the tools to access, retrieve, store, and evaluate, the plethora of existing medical information. In an attempt to integrate Medical Informatics into University of South Florida-College of Medicine curriculum, faculty designed a simple, interactive, user-friendly, problem-based www instructional delivery interface. A course specific template has been created that can be used for all the basic science courses, clinical clerkships and electives offered during the first four years of school. This framework serves as an entry point for access to each set of course materials. Course material consists of syllabus, resources, activities, and interactive practice tests. Our model provides better connections among faculty and students across different departments. It integrates subject areas into the curriculum that cut across many disciplines which are essential to the study and practice of medicine; it also integrates basic and clinical sciences to make the curriculum more clinically relevant. It provides the means to measure and analyze learning outcomes. This model places emphasis on students as active participants in the process of finding, organizing, analyzing, and applying information in novel ways to solve problems, communicate ideas, and continuously add to their knowledge base.

This presentation includes a real time demonstration of the curriculum template and "intranet", using multimedia, hyperlinks, and external web links. Anatomy, embryology, pathology, microbiology, family medicine, and medical informatics will be presented.

Developing Internet in Belarus: Minsk Internet Project

Sergei Kritsky <kritsky@unibel.by>
Nikolay Listopad <listopad@unibel.by>
Igor Tavgen <itavgen@bsf.minsk.by>

The main goal of the presented Project is to create a powerful Internet backbone network for the open society community in Belarus. The Project is supported by UN Development Program and Open Society Institute/Soros Foundation. It was started from Minsk Internet Project in December 1995. First results were reported at JENC'7 in Budapest (see JENC'7 Proceedings, pp. 174-1 - 174-4).

The tasks to be solved in the frame of the Project are:

- to set up a powerful IP backbone network in Minsk in order to develop Internet infrastructure around the capital (including telephone exchange nodes to interlink various local area networks), and provide it with high-speed Internet communications;
- - to make Internet access possible for a great number of organisations from the social sector in Minsk and Belarussian regional areas;
- to introduce and spread Internet culture and ideology as being a way of bringing together large groups of different users; and to carry out active educational and teaching programmes concerning computer networking for both users and specialists.

An Internet backbone network was put in place once the first stage of Minsk Internet Project had been implemented in Minsk. Fibre optic network is now connecting nine nodes including Belarusian State University, BELTELECOM headquarter, Centre of Information Security and UNIBEL Network Operation Centre located at Computer and Analytical Centre of Ministry of Education. Using the current equipment, it will be available for use by over 180 organisations and numerous private individuals in Minsk. At the moment, the backbone network is being used by 142 and various organisations from the Minsk social sector. Internet Training Centre has been established at Belarusian State University.

Currently we have got 256kbps line to BELPAK (official provider of Ministry of Communications). Fibre optical cable connecting Minsk backbone and BELPAK has been put in operation. This will give a possibility to use BELPAK's satellite connectivity to MCI (TELEPORT) and possible upgrade up to 512kbps.

The second stage of the project is to create Internet backbone nodes in Belarusian regional towns and hook them up to the Minsk backbone network. As a result of this stage, over 75 organisations from the Belarussian provinces will gain Internet access in the nearest future. Among them are Vitebsk State University, Vitebsk University of Technology, Polotsk

University, Gomel University, Gomel Polytechnical Institute, Mogiliov Technical Institute, Mogiliov Regional Library, Brest Polytechnic Institute, Grodno State University, Grodno Medical Institute and others.

The program's aims for future are to develop the infrastructure and increase the number of Internet users in the Republic of Belarus. We feel that the first priority for 1997 should be to provide international Internet connectivity. The second priority should be to develop infrastructure further in Minsk and the regions. In 1997-1998, we hope to hook up Belarusian higher educational institutions, institutes of the Academy of Science, etc. to the Minsk fibre-optic ring network. This will be achieved via fibre-optic channels and by installing routers to allow local area networks and various databases to be created. We feel that the third priority should be to develop our user network (assisting with connections to the existing network, and providing training and assistance to design specialized general-access databases for science, education, culture, legislation, etc.).

Collaborative Teaching in Cyberspace

Thérèse Laferrière,
Department of Didactics, Psychopedagogy and Educational Technology,
Laval University, Quebec, Canada, tlaf@fse.ulaval.ca

The design of a virtual community of communication and support for preservice teachers is an action-research project of the TeleLearning Network of Centres of Excellence (TL-NCE). Guided by the vision of interconnected learning communities (SchoolNet Canada), teacher educators from four Canadian universities (Laval, McGill, OISE/UT, UBC) are collaborating in learning how to prepare teachers for networked learning environments. Working with student teachers, teachers, and high-school learners, they use collaborative telelearning technologies such as Virtual U (http://virtual-u.cs.sfu.ca/vuweb) and WebCsile (http://csile.oise.utoronto.ca). The sociocultural barriers that have kept teachers isolated from one another are addressed using the research team model. They work together at building a repository of knowledge for teaching in cyberspace (http://www.tact.fse.ulaval.ca), one that emphasizes collaborative learning and teaching. To belong and contribute to one or a few computer-supported collaborative learning project(s) is seen as now critical for teachers.

Multimedia Presentations in Life Sciences Teaching

Revital Lavy, The Open University of Israel, Max Rowe
Educational center, 16 Klausner St. Ramat Aviv, P.O. Box 39328,
Tel-Aviv, 61392, Israel, revital@oumail.openu.ac.il.

Nurit Wengier, The Open University of Israel, Max Rowe
Educational center, 16 Klausner St. Ramat Aviv, P.O. Box 39328,
Tel-Aviv, 61392, Israel, nuritwe@oumail.openu.ac.il.

Booki Kimchi, The Open University of Israel, Max Rowe
Educational center, 16 Klausner St. Ramat Aviv, P.O. Box 39328,
Tel-Aviv, 61392, Israel, Bookiki@oumail.opneu.ac.il

Rina Ben-Yaacov, The Open University of Israel, Max Rowe
Educational center, 16 Klausner St. Ramat Aviv, P.O. Box 39328,
Tel-Aviv, 61392, Israel,
Rina_ben_yaacov_at_aft_ad1#@mail.icomverse.com

Advanced information and communication change teaching and learning methods continuously. Multimedia presentations are very potent in teaching biology, in facilitating spatial outlook and intuitive understanding: In visualization of complex structures; In simulation of complicated processes and mechanisms; In presenting laboratory instrumentations and techniques that generally are not available in students' laboratories; In exhibiting classic and complex experiments.

During university studies, particularly in distant learning, students confront problems understanding complicated mechanisms and enigmatic processes. The Open University of Israel, specializes in distant education, uses new methods and sophisticated techniques in order to improve teaching and enhance understanding levels among students.

We chose three main issues in biology, in order to demonstrate the role and importance of multimedia presentation in teaching Biology.

- The kidney: Illustrates the anatomy and physiology of the urine system.
- The nervous system: displays structure and function of the nervous system, and techniques used in electrophysiology research.
- Principles of Molecular biology: presents theory principles and techniques used in molecular biology research.

Acknowledgments:

We thank Gila Baron, and Chen Limor for graphic design.

Mentoring an Internet-based Distance Education Course Problems, Pitfalls, and Solutions

Marcia L. Marcolini, Clinical Associate
West Virginia K-12 RuralNet Project
West Virginia University
609 Allen Hall, PO Box 6122
Morgantown, WV 26506-6122 USA
mmarcoli@wvu.edu

LeeAnn Hill, Graduate Assistant
West Virginia K-12 RuralNet Project
West Virginia University
609 Allen Hall, PO Box 6122
Morgantown, WV 26506-6122 USA
lhill@wvu.edu

Increasing demands for advanced training is leading to an expansion of distance learning techniques in higher education. The West Virginia K-12 RuralNet Project, a National Science Foundation funded initiative, is a year-long training program providing teachers with skills in utilizing the Internet to enhance science and mathematics instruction.

During the Fall of 1996, on-line mentoring was utilized to facilitate completion of a graduate level distance education course completed by 122 RuralNet Teachers. Participants survey responses to perceptions of mentor contact, mentor assistance, perceived problems and suggestions for improvements indicate a mixed view of on-line mentoring.

While mentors were viewed a helpful especially in the areas of pedagogy and content, there were a variety of pitfalls including technical problems, lack of face-to-face interaction, contact initiation, timeliness and detail of response. Implemented solutions to alleviate such problems include a mentoring guide, mentoring workshops, mentor assignment considerations, and principles of adult mentoring scale.

The CREN Virtual Seminar Series: Learning at Your Desktop

Gregory A. Marks
Merit Network, Inc.
Ann Arbor, Michigan, USA
Email gmarks@merit.edu

Susan Gardner
Gardner Communications
Ann Arbor, Michigan, USA
Email gardnercom@aol.com

Rick Witten
Synapsys Media Network, Inc.
Ann Arbor, Michigan, USA
Email synaps29@idt.net

Suppose you already have experts leading high-quality, in-person presentations and discussions; can a similar experience be achieved via the Web? Can the production of such materials be done quickly and economically? If you want each learner in your audience to have control of the pace and path through the materials being delivered, can this be done with streaming video and audio? Yes, it is possible to accomplish these objectives; we are delivering many hours of advanced material about campus networking and Internet applications to staff at the hundreds of colleges and universities that are members of CREN, the Corporation for Research and Educational Networking. The presentations are delivered via streaming video and audio, with synchronized overheads, graphics, and other multimedia in well-defined Web browser frames, plus navigational controls and links to additional materials. Live discussions among participants are regularly scheduled events. This session will discuss our methods and their generalization to other content areas.

A Web-based Course in English as a Second Language: A Case Study

Ian Marquis
Language Instruction for Newcomers to Canada
North York Board of Education
Canada
nykenton@interlog.com

Jean Wang
Department of Linguistics
Simon Fraser University
Canada
jxwang@sfu.ca

Tài Nguyen
Language Instruction for Newcomers to Canada
North York Board of Education
Canada
thuan@accessv.com

This case study described the development, implementation and evaluation of a four week pilot Internet LINC (Language Instruction for Newcomers to Canada) course for immigrants.

The Virtual-University (V-U), a Web-based collaborative learning environment provided for asynchronous conference communication and for accessing a multimedia story, the focus of the students' course work. In addition, synchronous Internet Relay Chat (IRC), telephony software, and an electronic mail system were used in the course.

Data collected for analysis included V-U conference messages, IRC transcripts, questionnaires, and taped-interview transcripts.

Results showed benefits in promoting communication and language learning among students with different cultural backgrounds in a friendly learning environment very much like a real ESL classroom. Students' attitudes toward the on-line course were consistently enthusiastic and positive. The results also indicated that basic computer skills, training to use the software, and support from the instructors, were important factors for successful participation in the program.

Benjamin Franklin House: An illustration of a site management and visual design tool for complex, multi-authored web sites

Gil E. Marsden, g.e.marsden@mdx.ac.uk
Gareth J. Palmer, g.palmer@mdx.ac.uk
Harold Thimbleby, h.thimbleby@mdx.ac.uk
School of Computing Science, Middlesex University, Bounds Green Road, London, England, N11 2NQ

This poster uses the Benjamin Franklin Web Site, created for the Royal Society of Arts, to illustrate our research into distributed web authoring. Although distributed authoring raises many issues, we concentrate here on visual design. The site was created using a prototype tool named Siteview, which aims to promote consistency of design where authoring is done by many people. By separating page layout, structure and content, the various tasks of site creation can be assigned independently. Writers can concentrate on the content of their page, knowing that design elements such as navigation bars will be added later. Output pages are generated during a compilation phase before publication; this gives many of the advantages of dynamic page creation, but is more economical where pages are accessed more frequently than updated. Our extended abstract has further details.

Dmitry Sh. Matros

Distance Education Based on Computer Textbooks.

The contents of education taught at school or at University (it can be a textbook, a course of lectures, a manuscript and so on; to make it simpler we'll call it "a textbook" further) is put into the computer in the form of structural formulas, which are formed in the following way.

In the text of the book structural unities were singled out. For example, definitions, problems, questions, hypothesises, examples, principles, theorems and so on. Each structural unity is marked by a certain geometrical figure. Inside the structural unity its name is written.

Each structural unity gets its own number consisting of 3 numbers divided by means of points. The first number is the number of the chapter where the named structural unity is represented, the second number is the number of the paragraph, the third is the ordinal number of the structural unity inside the paragraph (apriory we'll suppose that any textbook consists of chapters and any chapter - of paragraphs).

Then the connections between structural unities are established. If the connection takes place within one paragraph, it is represented as a line consisting of horizontal and vertical cuts going from the earlier brought-in structural unity to the later one.

If the connection between structural unities takes place within different paragraphs, it is represented in the form of references. To the left from the structural unities are placed the numbers of the structural unities used in reproducing this structural unity, and to the right are enumerated the numbers of all the structural unities at reproducing of which this structural unity is used.

Hence, the user gets the whole information about the structural unity: its full name, contents, demonstration (if it has got any) and picture. Therefore, all the basic sentenses of the textbook appear on the screen, which together with the logical connections turns the electronic model into a teaching system. A "genealogy" of the structural unity is built, i.e. a chain of methods showing the line of methods used in the textbook which led to the structural unity. This information is extremely actual when Methods of Teaching Association is at work and when revision is being organised (besides, there is a special regime "Revision" for pupils).

When using the regime "Testing" check of pupil's knowledge on the paragraph, where the structural unity is situated, takes place. After inserting the user's name into the system, there appear on the screen testing tasks for every structural unity of the paragraph and every logical connection begot by this paragraph. On finishing the testing the information is memorized and on teacher's inquiry the result of each pupil is told.

This regime turns an electronic model into a controlling system.

The electronic model gives the teacher an apportunity to create a test for controlling for any paragraph of the textbook which is later used in the regime "Testing". Let's mark the principal differences between our programme of creating the tests and other available ones:

1) creating of tests takes place in the regime of the dialogue between the computer and its user;

2) computer is making the user go logically through the structural formula of the chosen paragraph, helping in creating tests for every structural unity and every logical connection; it garantees validity, system and systematic character of the recieved test;

3) part of the tests is created by the computer automatically without any interference on the user's part but following his instructions;

4) others are created computer-aidedly jointly by the computer and its user;

5) a ready test can be used in any of the following regimes:

a) check of knowledge while looking through the textbook

b) autonomous check of knowledge with the help of the computer

c) producing the prepared test to carry out the checking

d) according to the theory of testing 4 forms of testing tasks are created: closed, open, tests to find out accordance and knowledge of the succession.

It's evident that electronic model is open for further development (and so it happens) so any teaching programme is sure to be based on the corresponding contents of education.

Such models are imparted to those taught by TV means of communication. They are used both as teaching and illustrating systems. The students study the material successively moving from one paragraph to another. Doing it the operative back communication allows to correct the process of teaching in time.

The correction is necessary because the system tells a student his situation in the structural formula at any moment, i.e. what structural unities have already been adopted by him, what have not and tells him his main mistakes. Following this data the rating of those being taught and pedagogical monitoring are built up.

The discribed models are built and realized on many subjects: language, Mathematics, History, Physics and so on.

Political Philosophy and the Technology Curriculum

Bruce W McMillan
University of Otago
New Zealand

bruce.mcmillan@stonebow.otago.ac.nz

New Zealand has experienced several years of social, political, and economic reforms, which emphasised internationalism and competition as the force for educational achievement. This has pressured schools to prepare students for work, and brought about a changed relation between teachers and community. Technology is seen to play a central role.

Without a careful analysis of the links between the technology curriculum, teaching practices regarding technology, and the wider social context for using technology, the promised benefits to education may be illusory. Technology is playing a significant role in the deconstruction of the welfare state, and the establishment of a radical monetarist economy. Thus education for "the information age" involves a different conception of society. That is a debate which cannot be left to the information technologists alone.

A paper is available from the author.

What it Really Takes to Put Your Lab on the Internet

John Mertes, Director of Technology
Rhodes School, USA
email: jmertes@rhodes.k12.il.us

Schools have choices when it comes to going on line. If those choices are not correct, problems arise that are both expensive and frustrating. Whether it means installing new equipment, or using what you already have, a solution exists for both cases. We basically want to share our experiences to benefit those in the same situation. The pros, the cons, as well as what we would do differently will be discussed. Using the internet we will show our homepage which includes a pictorial tour of our hardware installed.

Additionally, what makes our site somewhat special is the partnership we developed with our local college, Triton College. Via wireless technology we have connected with them as well as sharing the cost of this connection. A winning case for all.

Web-based Education: Considerations and Approaches

Jay Moonah
Digital Media Projects Office
Ryerson Polytechnic University
Toronto, Ontario, Canada
jmoonah@acs.ryerson.ca

As an institute with a rich background in the areas of media production, as well as computer and telecommunications technologies, Ryerson Polytechnic University has made a major commitment to pursuing the educational possibilities presented by the use of the World Wide Web as delivery medium. A number major projects have already been developed, and many more are in the works. Along the way, Ryerson has gained invaluable experience in the areas of effective conversion of existing material to a Web-based format, authoring for multiple delivery environments, using the Web as a medium for teacher/student interaction and providing faculty support and training in multimedia and Web technologies.

Demonstrations will include large scale projects (Interactive Learning Connection/University Space Network, the Eaton School of Retailing and the CourseVault pilot projects,) courseware initiatives (the BIA Insolvency Counsellor's Qualification Course and Digital Applications MPS024,) and faculty support modes (the Digital Media Projects Office.)

Multi Media - Malaysian National Curriculum

Vijaya Kumaran K.K. Nair
Manager
Database and Research Center,
Stamford College Berhad
NO. 17 & 19, 3rd Floor, Wisma Yan,
Jalan Selangor,
40650 Petaling Jaya,
Selangor Darul Ehsan.
MALAYSIA

Tel: 03 - 7551300/7572300
Fax: 03 - 7551310
E-Mail: vjn@pc.jaring.my

The Pusat Sumber Ilmu (PSi), is an on line database designed for students. It is a step taken by Rangkaian Tenaga Sdn Bhd (RTSB) to create opportunities for information accessibility. The PSi relays knowledge to schools nationwide, providing students easy access to information on any subject within and beyond the school curriculum. The system offers a multi media based information (featuring text, full color graphics, sound, nimation and motion video capabilities) on network computing. It also provides a wide range of educational database that a subscriber can access via computers hooked onto the PSi network.

The Stamford Database & Research Center is committed to developing the above on line library of information encompassing the whole spectrum of studies at the National Secondary School level. This is termed as the Latihan Kurikulum(curriculum exercises). Databases have been created for all the PMR, SPM and STPM subjects on a multi media platform. The compilation of subject - specific material has been undertaken by more than 100 senior educationist and reviewed by qualified professionals based in tertiary institutions. In addition to the approved curriculum, this project would also have other features such as :

- Examination Format
- Comprehensive Subjects Resources

HOW TO PROVIDE PUBLIC ACCESS TO INTERNET INFORMATION SOURCES ON PUBLIC ACCESS WORKSTATIONS?

Paul Nieuwenhuysen
Vrije Universiteit Brussel,
Pleinlaan 2,
B-1050 Brussels - Belgium
Tel: ++32-2-629.24.36
Fax: ++32-2-629.26.93
pnieuwen@vub.ac.be

More and more useful information becomes available online through the Internet, accessible by using only one integrated package of Internet client programs, with common and affordable hardware. Much of this information is free of charge. Therefore, offering public access to this information becomes feasible in libraries, schools, and similar environments.

This contribution points out interrelated problems, questions and options related to client hardware, server computers, data communications, client software, personal disk space for users, security risks, the scattering of information, marketing of the service, electronic mail facilities, "free or fee", personnel, and user guidance. Acceptable solutions and answers depend of course on the environment. The overview can serve as a check list. At least, it shows that many options exist and that offering and optimizing public access is not straightforward.

Global Educational Database on The WWW(World-Wide Web) and Its Application in School

Masatoyo OHSHIMA
Kanzaki-Seimei Senior High School, Japan
ohshima@saga-ed.go.jp

Yasuhisa OKAZAKI
Department of Information Science, Saga University, Japan
okaz@ai.is.saga-u.ac.jp

Hiroshi NOKITA
Yamato Junior High School, Japan
nokita@saga-ed.go.jp

Hidekatsu HARA
Saga Prefectural Government, Japan
hara@saga-ed.go.jp

Hisaharu TANAKA
Department of Information Science, Saga University, Japan
kangaroo@ai.is.saga-u.ac.jp

Hirofumi Momii
Saga Prefectural Education Center
momii@saga-ed.go.jp

Kenzi WATANABE
Department of Computer and Communication Sciences, Wakayama University, Japna
watanabe@sys.wakayama-u.ac.jp

Hiroki KONDO
Department of Information Science, Saga University, Japan
kondo_h@ai.is.saga-u.ac.jp

Although the WWW has huge amount of information, teachers have found that it is difficult to reuse much useful information on the WWW for classroom lessons. The reason is that databases on the WWW are constructed based on a designer's own viewpoint. The problem is that it does not always match the context of the class.

We introduce "link information" as an interface to quote information, whose original viewpoint is different. It changes the view from original one to user side one. We present an architecture of building a WWW database, which utilizes rich information on the WWW by virtue of "link information." We call it as "User Side Database," because it stands on the viewpoint of its user.

We apply this framework to an educational WWW database as a teaching material for classroom lessons. We have builta User Side Database for environmental problems [Nokita 96] and investigated its effectiveness through classroom lessons in a Japanese junior high school.

References

[Nokita 96] H. Nokita(1996). A Room for Learning Environmental Problems. http://www.saga-ed.go.jp/materials/info.kankyo.html

Canada's Wired Writers: The Writers In Electronic Residence Program

Trevor Owen
Faculty of Education, York University, North York, Ontario, Canada M3J 1P3
Co-ordinator, Instructional Technology/Online Learning in Teacher Education.
Program Director, Writers In Electronic Residence
Web site: www.wier.yorku.ca/WIERhome/
Email: wier@edu.yorku.ca

ABSTRACT

Writers In Electronic Residence (WIER) links Canada's writers with Canada's schools in an exchange of original writing and commentary. Well-known authors join classrooms electronically to read and consider the students' work, offer reactions and ideas, and guide discussions between all participants.

Like familiar "writer in (conventional) residence" programs, WIER brings writers into classrooms. Unlike more conventional residencies, WIER's "residencies" are undertaken in an in an online computer conferencing environment. Participating schools also receive copies of books written by the authors with whom they work online.

WIER is a national program, offering 12 week programs at all grade levels each fall, winter and spring. WIER's web site offers program information, writer biographies, student writing samples, as well as resources for writing and educators.

SELENA: Walking on the Moon or How to Make Decisions Using the Web

VALERY A. PETRUSHIN, MARK YOHANNAN, AND TETYANA LYSYUK
petr@cc.gatech.edu
Georgia Institute of Technology
Atlanta, GA

SELENA is a Web-based decision support system which was initially developed to support selection stages of design process for Georgia Tech's mechanical engineering students of the class ME3110: Creative Decisions and Design. But it can be used in any design class by students of various schools. It also could be found helpful for professional designers. The name "selena" on the one hand is close to the word "selection", but on the other hand it originates from Greek word "selene" which means the moon. As you know the lunar gravity is six times less than the terrestrial gravity. This means it is easier to walk on the moon than on earth, but you must be equipped. SELENA just provides the tools (equipment) which allow students and professionals walk easier through the selection stage of design.

ME3110 Class

Third-year mechanical engineering students at Georgia Tech are currently required to take the class ME3110. The students in this class are introduced to a specific design framework: the Decision Support Problem Technique that is based on a Decision-Based Design approach [Muster and Mistree, 1989; Mistree et al., 1993]. Though the course covers both meta-design (planning, scheduling, reporting) and design activities (preliminary selection, selection, compromise), the main focus is teaching the students to partition the problem into subsystems and select among concepts to meet the functional requirements for each subsystem. During the course of a ten-week quarter, students form teams, design and build a required product which is a mechanical device that solve the problem, present it both in a demonstrative competition and a sales presentation to customers who are their classmates and the professors of the course. The students have found the course both exciting and challenging. More information about the ME3110 course is available on the Web at http://srl.marc.gatech.edu/education/ME3110/me3110-Web.html.

SELENA's Objectives

According to the Decision Support Problem Technique the selection consists of two major phases: preliminary selection for identification of a set of potentially superior concepts based on qualitative rather than quantitative information, and selection for identification one or a very limited number of superior alternatives among the concepts selected earlier, using both insight-based "soft" information and science-based "hard" data. The main objectives for implementing SELENA are to create a Web-based tool which:

- Seamlessly supports both phases of selection process.
- Helps students to create design reports.
- Contains conceptual information and examples to support performance and facilitate learning.

Implementation

SELENA uses the Analytical Hierarchy Process for decision making [Saaty, 1982]. It was implemented on an Apache UNIX server, using HTML, JavaScript, and PERL. It is available to public at http://srl.marc.gatech.edu/education/ME3110/selena/Selena.html.

Acknowledgments

We are grateful for the continuing support and encouragement of Farrokh Mistree, Janet K. Allen, and J. Chris Thompson. Research reported here has been supported by NSF grant DUE-945388, and by the Woodruff Foundation's support of the EduTech Institute.

References

Muster, D. and Mistree, F., 1989 The Decision Support Problem Technique in Engineering Design, *International Journal of Applied Engineering Education*, 4:1, pp. 23-33.

Mistree, F., Hughes, O.F., and Bras, B.A., 1993 The Compromise Decision Support Problem and the Adaptive Linear Programming Algorithm. In M.P. Kamat (ed.) *Structural Optimization, Status and Promise*, AIAA, Washington, D.C., pp. 247-286.

Saaty, T.L. 1982 Decision Making for Leaders. Lifetime Learning Publications. Wadsworth, Inc., Belmont, CA.

ISSUES OF AUTHORITY IN ON-LINE INSTRUCTION

JoAnne Podis

As professors move into electronic settings, the various material qualities of their interactions with students, many of which combine in the classroom to become seats of their authority as professors, disintegrate, with potentially major implications for those of us who contemplate on-line instruction.

In this study I explore possible responses to questions such as the following: From what sources does our authority as professors tend to derive? How do those sources change in an electronic setting? How do the students' educational/social/personal contexts influence the authority relationship on-line as opposed to within the classroom? And finally, does the authority dynamic between professor and student change by design or of necessity? If the latter, what are the implications for professors contemplating such instruction? In addition to my own experiences my discussion is informed by the experiences of colleagues who have taught on-line.

Codes of Ethics for Computing at Colleges and Universities in the United States Revisited

Lester J. Pourciau
Director of Libraries
The University of Memphis
Memphis TN 38152, U.S.A.
Voice 901 678-2201
FAX 901 678-8218
Email pourciau@cc.memphis.edu

When American colleges and universities began, several decades ago, to make use of computer technology, the notion of "ethics" as applied to their efforts was, for the most part, an alien thought. One might assume, as colleges and universities began to make efforts to bring computer and information technology into full play in the enterprise of education, that commensurate efforts were made to establish Codes of Ethics for the use of this technology. The effort reported by the original study under discussion here was a thorough examination, in the spring of 1996, of fifty American college or university Home Pages, selected at random, for evidence of promulgated codes of ethics for computer use. Twenty-six Home pages were found to provide links to such Codes. The effort reported here is a reexamination of the remaining twenty-four Home Pages for evidence of such links.

Putting Large Volumes of Information on an Intranet

Ann Rockley
President, The Rockley Group Inc.
Canada
rockley@rockley.com
http://www.rockley.com

Intranets are becoming common in corporations. Intranets allow corporations to distribute information in an effective way throughout the organization. Intranets take advantage of web technology which provides a fast and cheap solution for information distribution. However, intranets often push this technology to the limits through the use of large volumes of information and multiple authors throughout the organization. This paper addresses the issues of analysis, design, tool selection and management of large volumes of information on an intranet.

All too often large documents are "dumped" online. No changes are made to the paper materials to accommodate the new media (change page size, organization, structure, etc.). This may seem the fastest and most cost-effective route; however, the costs incurred by users trying to use this information, or often not using this information because it is too frustrating, can far outweigh the early cost savings. Online is a very different medium than paper and just as you wouldn't present information on video as you would on paper, you shouldn't present information online in the same way as you present paper information. To determine the most effective ways to put your materials online begin with analysis, then create criteria for tool selection and finally design effective materials for use online.

Analysis

Analysis should consist of audience, information, authoring and maintenance, and hardware and software analysis. From this analysis you can create a criteria list for selecting an appropriate tool, and designing effective materials.

Audience Analysis

The audience is analyzed to determine how their characteristics and needs affect how to put the documents online. Understanding the audience will allow you to determine content, organization, breadth, depth, access methods, and presentation methods. Corporations sometimes assume that it is not necessary to conduct an audience analysis of internal staff because their characteristics are well known; however, their profile must be revisited to review how their characteristics will affect the design of effective online materials.

Information Analysis

Information is analyzed to determine how effectively it will go online. Different types of information work best presented in a particular way. You need to review the materials to determine:

how well it is written (long passages of text do not work well online, short chunks of information are better)
if materials are consistent both in look and feel and writing style (affects conversion and usability)
how tables are used (difficult to put online, and difficult for users to use large tables online)
how graphics are used (graphics which do not display well online are often not worth including)
the relationship between information (cross-references, levels of detail, implied links to other

sections/documents)

Authoring and Maintenance Analysis

The people who originally authored the information and those that are likely to continue to maintain the information are important to the tool selection. Determine:

if they would be comfortable authoring in something like HTML or would they prefer to work in something like Word or Wordperfect

if the authors are familiar with concepts of Internet materials and other online documentation (in order to assist in the re-engineering process or will this be a steep learning curve/will someone else be required to assist)

if there are multiple authors

if there currently is a quality assurance process in place to ensure effective control of documents as they are authored almost instantaneously

if workflow and revision control are required

Hardware and Software Analysis

Review the existing hardware and software in-house to confirm that they are adequate to sustain the requirements of authoring and managing the online materials. Also review your customer/user hardware and software to determine if they will be capable of displaying the materials once created. We have found that it is not unusual to have a large number of "dumb" terminals or low-end PC's in large corporations.

Tool selection

Selecting the right tool to create and manage the information is very important to the success of the project. There are pros and cons to every tool. The nature of the information will often dictate the tool. The following provides some insight into the pros and cons of some standard formats.

HTML vs. Acrobat

HTML is the defacto standard for the Internet and now intranets. Acrobat is an alternate Internet standard to HTML. It is what is known as a portable document format, which allows you to create a viewable file that looks identical to the paper version.

HTML		Acrobat	
Pros	**Cons**	**Pros**	**Cons**
Small files that are very fast to access and distribute	Layout (look and feel) capabilities are limited.	Fast conversion of legacy documents (can produce an online document that looks identical to the paper document very quickly)	Documents designed for paper are not effective for use online. Therefore documents that have been converted to Acrobat exactly as they appear on paper are best suited to printing, not to use online.
Most used format on the Internet/intranet	Deals poorly with layered documentation	Excellent for materials where the visual presentation is	Large to very large files produced which make access and use very

		important (i.e., brochures, newsletters)	slow.
Largest focus of third party software solution providers	Poor navigational capabilities (e.g., contents, index must be manually created)	Excellent for display of graphics (user can zoom in for detailed viewing)	Interactive functionality of the document is limited to basic links.
Functionality can be enhanced by Java.	Tools provided to author in HTML are limited in comparison to standard word processors	Easy display and creation of table of contents	Deals poorly with layered documentation
	Minimal functionality (e.g., tables of contents and indexes must be manually generated, layout capabilities are "primitive", searching must be added in).		Minimal functionality (e.g., only one table of contents allowed, indexes must be manually generated, can only search within a document, not across documents).

Table 1: HTML vs. Acrobat

HTML vs. SGML

HTML can be thought of as a very small subset of SGML (Standard General Markup Language). SGML was developed in the early 70's by Charles Goldfarb (IBM) as an outgrowth of DCF/IPF (both document tagging systems) to provide a standard for defining documents. It was hailed by large industry (military, aerospace, telecommunications, government) as a solution to their problems of multiple incompatible document formats and multiple platform problems. It provides considerable power to online documentation.

HTML (Hypertext Markup Language) is also an outgrowth of DCF/IPF but it was developed specifically for the Internet. It is a much smaller tagging language than SGML which makes it easier to learn, but less powerful to use.

SGML	
Pros	**Cons**
Platform independent	Steep learning curve
Describes the content of information not just the format so that information can be retrieved based on content, not just text.	Expensive to design and develop a DTD (Document Type Definition) necessary for the effective use of materials.
Powerful database capabilities	Very expensive software to create and manage ($10,000US to $25,000US).
Powerful information reusability capabilities	If you want to use SGML on your intranet you must use a specific SGML browser rather than an inexpensive SGML browser
	Must be "dumbed-down" to run as HTML on the Internet/intranet so

	that much of the added functionality is lost.

Table 2: Pros and Cons of SGML

"On the fly" Conversion to HTML vs. native HTML

There are now some tools which provide "on-the-fly" translation of the source information into HTML. Lotus Notes is the most popular of these tools. FolioViews also provides this capability. These tools also provide powerful workgroup capabilities and basic document management.

"On-the-fly" Conversion to HTML	
Pros	**Cons**
Good word processing capabilities	There is a slight delay in user receipt of information
Basic document management capabilities	Need to do additional work to map the "converted" materials to an attractive HTML format
Good link management	Link names as displayed in the user's browser can be totally incomprehensible
Automated TOCs and index	Native tool is more expensive to purchase than an HTML authoring tool
Excellent integrated search facilities	

Table 3: "On-the-fly" Conversion to HTML Pros and Cons

Managing your information

Large paper documents can be difficult to manage and control, but large online documents can be a nightmare if you do not use document management software from the beginning. There are many different types of ways you can approach managing your materials.

Integrated Internet Development Systems

Integrated development systems for building, publishing and maintaining web applications offer many tools that are not available to companies who built intranets from HTML-coded static pages. These environments aim to handle all aspects of web creation, including application development, content creation and page layout. Proprietary binary formats eliminate the common problems of broken links. Developers can create relationships among web application objects so that when a linked page is moved, the tool will fix the link. These tools have a higher cost than the lower-priced tools aimed mainly at content creators. The higher price reflects the programming support and sophisticated debugging environment that can be used to create commercial website with interactivity and multimedia.

Site management software

Site management software has industrial-strength tools that cost more and do more than simple Web authoring tools. These tools specialize in managing the links as pages are updated, making it easier to move web sites from one place to another. There are three different types of site-management software:

1. site management software included in an integrated development environment
2. stand-alone site management software such as Build-it
3. site management software combined with authoring

Site management software in integrated development environments is sometimes powerful enough to allow whole web sites to be moved by clicking a button (even to a different operating system) but creating content in an IDE is still laborious and time-consuming.

Stand-alone site management is most useful for commercial web-sites. Stand-alone products such as Build-it offer site management and a controlled development environment for programmers and content creators by integrating the site management toolset with a third-party software source tool. This type of tool does not address the problem of making content creation faster and easier.

Combination site management and authoring tools do not have the power in site management that is offered in the tools for IDE and stand-alone site management. This is because authoring tools are essentially a step-up from coding static pages by hand with the higher-end ones adding some interactivity for forms and automated addition of repeated elements like navigation bars or copyright notices.

Workflow applications

Workflow is available in many different "flavors" including Ad Hoc, Object-Oriented, Transaction Base, Knowledge Base, and many others. Currently the most popular types of workflow are Ad Hoc, Object Oriented, and Transaction based. Workflow applications are typically client-server but many of these products have add-ons that allow the workflow to be web-enabled.

Ad Hoc is designed for processes that must be handled on a case-by-case basis. It allows routing to be mapped graphically, monitored and changed as needed. Visual Mapping helps people involved in a process to see where they fit. Ad Hoc workflow is a good solution for workgroups or departments that must deal with rapidly changing environments.

Object Oriented workflow utilizes predefined objects as the underlying architecture. This is a good solution for companies with processes that can be defined using a common set of components that may need to be reorganized on a regular basis. The processes behind the object are never manipulated by the end-user. Certain objects will fit with other objects so that workgroups or teams can create processes that will conform to company policy, eliminating the need for formal approval. This application could be used for creating workflow for authoring, editing, approving and distributing documentation.

Transaction based workflow is similar to a chain reaction: for every step there is another step that leads to another step. Each step in the process is predefined and follows a certain route depending on actions and outcomes. The process continues with very little human intervention until it reaches completion. This application could be used in the areas of forms processing for loans.

Design

Designing documents for online often means re-engineering them. Some areas of design to consider are:

Modularity

Information should be chunked into smaller modules for:

easy access to information
manageable pieces of information
reusable information

Scannability

It is difficult to read large volumes of information online. Users tend to visually scan the text to pick out important pieces of information. Use:

lists instead of paragraphs
short tables or columnar presentation of information
white space (don't tightly pack information)
sub-headings to break up information
short paragraphs (3-6 sentences)
a consistent design for different types of information

Layer information

Information online should be short and precise. However, sometimes it is necessary to provide more information. You can layer information with one level presented and subsequent levels linked through secondary windows or pop-ups. Ensure that users know where they have come from and where they should go.

Hierarchy

Create no more than three levels of information. More than three levels of information is difficult to navigate, particularly without a hierarchical table of contents.

Provide continuity/connection

Chunks of information in an online environment are short and discrete. It is difficult to know what has gone before and what comes afterwards in continuing processes.

Indicate order through titles (e.g., number the titles).
Explicitly refer to preceding or following processes (e.g., "This is the second step in creating..." or through related topics)
Provide links to the processes (e.g., a bulleted list of the 1st, 3rd, 4th steps in the process which are linked)

Overviews

Create overviews to sections to provide context for what is to follow. For example:

This process consists of xx steps.
Label it "Overview" to give users the option of selecting it or not.
Provide cross-references (links) to the related sections.

Design standards

Create design standards and guidelines for use throughout the organization to ensure that information is consistently designed for easy access by the users. Different types of information may require different design standards, but there should be a core of consistency.

Retrieval

Your document(s) are only as good as the user's ability to retrieve information. Full-text retrieval is not enough. A combination of the following should be considered:

table of contents
traditional index
full-text retrieval

Conclusion

Putting large volumes of information online requires a great deal of analysis, design and often re-engineering of information. However, the payback in terms of instant access to current and accurate information can save an organization millions of dollars in previously lost productivity.

Web-supported learning by example.

Marco Ronchetti
Dipartimento di Informatica e Studi Aziendali
Universita' di Trento, 38100 TRENTO Italy

Over the last decade, we have seen a shift in the programming paradigm. Software development environments tend now to be very rich. Huge libraries (like the X-library, the Macintosh Toolbox etc.) offer a wide choice of functionality which relief the programmer from repeating common tasks.
However, in order to get advantage from these tools, the programmer must:
- Be aware of them,
- Know how to use them,
- Understand the relationships among items.

The learning barrier in these environments is very steep and high, and the advantages are available only after this barrier has been passed.

Object Oriented technology helps a bit by providing a hierarchy and by encapsulating features: the problem is however not completely solved.
Moreover, to achieve software reuse often means to build domain-specific classes: but in order to be "reused" these classes must in first place be known and understood.

The very popular Java environment also suffers from this syndrome. In fact, although one day can be enough to learn the language, a much longer time is needed to know to a decent level its class library.

The most effective technique for teaching any OO technique seems to be the "mentoring": i.e. one teacher guides (at most three) disciples through the secrets of the new knowledge. [AUE95]. Java is no exception. Such practice is however very expensive: therefore surrogate tool which allow self-teaching are sorely needed.

A typical approach to solve the problem is to offer a tutorial, which presents the most common classes, methods and examples of how-to-use them. Most Java books do exactly that, and many among them reach the goal. However, since new class packages are constantly being written to address specific needs (like electronic commerce, interface to databases, distributed objects, etc.) the programmer has to always learn new things. Moreover, such approach is not focussed on the specific needs of the user, but rather offer an "average" solution.

We are currently working to propose a solution that can help overcoming the problem by taking advantage of the Web capabilities. Our solution implements a "Web-based Software Repository" which allows archiving artifacts through the web. We start from the concept of a library for collecting reusable assets [SUC95]. In general such libraries are built to increment the level of reuse within an organization. Their role is to make known and available valuable artifacts to everybody. By artifact we mean a document related to any phase of the software life cycle: a piece of code, documentation, a specification document etc. Typically artifacts can be searched, retrieved and used. Relationships among artifacts are also supported.

We believe that such a "software repository" can be successfully used to support a more focussed "learn by example" paradigm, where the needed information is supplied on a "just in time" basis. A programmer needing to solve a particular problem could use a search mechanism based on keywords and free text search. A more refined mechanism based on faceted classification [PRI91] is available to the user to perform a more focussed search. In order to understand a particular class, s/he could use the repository to find examples that provide useful insight.

For instance, in the scenario we envision, a user wishing to create a client-server application could be led by the search mechanism into the "java.net" class hierarchy. There s/he could find the list of classes, their documentation, and examples of code using those classes. Each user can then grade the examples s/he used, leaving a track that could help other people to choose the best examples. The user can also annotate the available examples, leaving hints useful for others. The system could therefore evolve, improving its ability to help people.

We are currently in the process of implementing the server-side of a Web-based Software Repository. The client side of our system, written in Java, can run as an applet in any Web Browser. The server side, also written in Java, can run on any machine thanks to the intrinsic portability of the language.

We plan to finish the implementation soon, and to perform experiments our Web-based Software Repository during the next academic year with our students. An extension of the concept with the aid of multimedia, along the lines suggested in [FER96] is also being evaluated.

References

[AUE95] Auer K., "Smalltalk training: as innovative as the environment", Comm.ACM, 38, 10-115, 1995

[FER96] Ferreira M. And Werner C.M.L., "Packaging reusable components using patterns and hypermedia", in "IVth International Conference on Software Reuse", Sitaraman Ed., IEEE Computer Press 1996, p. 146.

[PRI91] Prieto-Diaz R.: "Implementing faceted classification for software Reuse", Comm.ACM, 34, 5-89 (1991)

[SUC95] Succi G., Ronchetti M., Uhrik C., Baruchelli F., Cardino G., Valerio A. "Gestione delle Configurazioni in Ambienti Multiorganizzazione Distribuiti usando Sarto." In Atti del Congresso Annuale AICA. Chia (CA), Italy, September 1995
Succi G., Uhrik C., Ronchetti M., Valerio A., Cardino G.,"The Role of a Configuration Management System into a Reuse Oriented Framework." International Journal on Applied Software Technology. 3/4 (1995), 237

Design consideration in the WEQ-Net site development

Katia Barbosa Saikoski, Ms.C.
katia@inf.pucrs.br
Computer Sciente Institute

Mara Lucia Fernandes Carneiro, Ms.C.
mara@eq.pucrs.br
Department of Chemical Engineering

PUCRS - Pontifical Catholic University of Rio Grande do Sul - Brazil
1997

The WEQ-Net project establishes the joint of two groups: Chemical Engineering Department and Computer Science Department / PUCRS to design and develop a site, called WEQ-Net, about Chemical Engineering, including theoretical information, software, databases, helping modules. To develop this site it was necessary to evaluate educational aspects related to the teaching and learning processes.

WEQ-Net Modules

Each module of the site is composed by three sub-modules:

- theory module: concepts related to chemical engineering.
- educational module: creation of an environment for supervised study in groups with a supervisor that creates, submits, corrects and discusses exercises.
- software module: development of software tools related with the theory module.

Project Steps

- identify the necessary elements to compose and implement the theory and software modules;
- implement the educational;
- create the homepages;
- implement or adapt the necessary software to the software module;
- test and validate the implementation.

World Wide Web Hypertext Linkage Patterns

Dr. Perry Schoon, Illinois State University, USA, pschoon@fau.campus.mci.net

The purpose of the study was to: (a) investigate the efficiency of navigating World Wide Web sites constructed using different hypertext linkage patterns; (b) identify the differences between experienced and inexperienced World Wide Web users in their efficiency in navigating web sites constructed using different hypertext linkage patterns; (c) identify the differences between males and females in their efficiency in navigating web sites constructed using different hypertext linkage patterns; and (d) to identify any interaction effects between gender and experience on the efficiency of using any of the linkage patterns.

Data were collected from 261 participants through demographic and experience questionnaires, activity sheets, and computer generated text files. Results of the analyses showed that web sites patterned after star and hierarchy linkage patterns were more efficient to navigate for informational use than were web sites patterned after linear and hierarchy linkage patterns. Females were shown to have a much more difficult time navigating arbitrary web sites than males.

ProMediWeb - Medical case training and evaluation using the World Wide Web

Hardy Schulze (1), Thomas Baehring (1), Martin Adler (3), Sepp Bruckmoser (2), Martin Fischer (3)

(1) Department of Internal Medicine, University Hospital of Leipzig, Germany
(2) Institute of Educational Psychology, University of Munich, Germany
(3) Medical Hospital, University of Munich, Germany

Contact:
Dipl.-Inform. Hardy Schulze
University Hospital of Leipzig
Department of Internal Medicine
Ph-Rosenthal-Str. 27
04103 Leipzig, Germany
Tel: (+49 341) 97 13 211
Fax: (+49 341) 97 13 259
E-Mail: shuh@server3.medizin.uni-leipzig.de

The present alterations in the curriculum of medical education emphasize particularly on problem based learning strategies which only can be trained on close-to-reality cases. In this way the future physicians can gain knowledge and skills which they need in their later carrier to handle real clinical situations. Software applications, which give the opportunity to work on authentic-designed clinical cases and therefore support this learning process. In Germany alone, there are 60,000 potential users, students of medicine in the clinical terms.

A comprehensive integration of computer-aided learning programs into the curriculum has failed so far because of insufficient technical availability and non-crossplatform applications. The World Wide Web (WWW) provides a suitable platform for distribution and easy handling of medical teaching software. In Germany the reinforcement of the German Research Network allows a high transfer capacity for multimedia data from the resource WWW server to the student application. Furthermore it gives the possibility to establish cooperative learning strategies beyond the borders of subjects or universities and to integrate those into the medical education.

In this time the medical education tends to a more problem and case based approach. Therefore our objective is to offer realistically designed medical learning cases in an interactive way by using the World Wide Web. The ProMediWeb system is dedicated to learners and authors of medical cases as a standard-WWW-application. It is developed by computer scientists, physicians and psychologists. Using a data window and a control&communication window the ProMediWeb server carries out the selection, presentation and interactive engagement in the medical case via the HTML standard. Special play, cooperative and communicative servers are developed to provide those functions. A cooperative setting between users and the learning case is provided by an on-line communication (hierarchic chat function). Furthermore a dedicated cooperative engagement of 2 learners dealing with the cases is possible. Case-related comments can be retrieved via case newsgroups. Because of the interactive design of the ProMediWeb learning system an only passive and isolated learning of the students is avoided; they are able to pass an Internet-based dialogue with other students and even with the case author to solve the problem in an active way.

The ProMediWeb application is realized with a specially designed software using CGI and client applet technology (MS Visual C++©, Java, JavaScript©). For communication purposes we have integrated standard tools such as case newsgroups, chat and audioconferencing. The multimedia case material is stored in an object-oriented database (NeoAccess©) at our WWW teaching servers in Leipzig and Munich.

Pre-use and post-use questionnaires (HTML design) and an user and interaction database on our WWW-Server allow an evaluation of learning behavior and acceptance from the side of the users. Simultaneously we will start an evaluation of the system as a part of the practical training "medical teaching software" in the curriculum of the University of Leipzig. Motivation, quality assessment and acceptance of about 400 medical students will be registered and analyzed.

This poster will present the didactic and technical structure as well as the concept of the practical evaluation of the ProMediWeb system. The software may be demonstrated in its first stage.

Risk Assessment and Training about Type-2 Diabetes on the Internet

Hardy Schulze (1), Thomas Baehring (1), Stefan R. Bornstein (1), Werner A. Scherbaum (2)

(1) Department of Internal Medicine, University Hospital of Leipzig, Germany
(2) Diabetes Research Institute c/o University of Düsseldorf, Germany

Contact:
Dipl.-Inform. Hardy Schulze
University Hospital of Leipzig
Department of Internal Medicine
Ph-Rosenthal-Str. 27
04103 Leipzig, Germany
Tel: (+49 341) 97 13 211
Fax: (+49 341) 97 13 259
E-Mail: shuh@server3.medizin.uni-leipzig.de

More than over 100 million people suffer from the glucose-deficiency disease Diabetes mellitus. This number is expected to double over the next 15 years due to further changes in lifestyles and the increase in life-expectancy.
Diabetes mellitus is the result of a complex metabolic disease, which can lead to acute problems and extreme consequences. Characteristically, there is an increase in the blood-sugar level. The number of undiagnosed cases is estimated to be between 20% and 50%. Through the medium of a simple questionnaire on the WWW, a pre-selection of individuals with the possibility of an increased risk-potential of Type-2 diabetes could be examined.

A test with 79% accuracy for Diabetic-prevalence 5% [1] was adapted for us through the WWW.

We went back to the HTML Frame-Technology format established by the firm Netscape in order to present the user with an interesting overview Front-End.

It should be noted, however, that access to the test is possible with all of the WWW-Browsers. Also, so that the test can be used worldwide, it has been formulated in both German and English.

After calling up the URL, the user is asked to enter his/her data to be assessed into the computer. The data includes such characteristics as age, weight, and height. After entering this data, it is sent via Internet to our WWW-Server. This activates the CGI-program which analyzes the corresponding data and saves it confidentially in a database. Also included is such additional information as IP-address and Hostname.

The CGI-program generates a dynamic HTML page which includes the results of the risk-test (the potential increased risk that exists for the user to contract Type-2 diabetes) and several statistical output (number of already completed tests, comparison of personal weight with the mean of all test-takers, etc.).

This page is sent to the WWW-Server and from there then sent back to the respondant's browser. Through the Diabetes Risk-Test Homepage additional information on the disease can be called up. Also offered are links to national and international organizations and discussion forums.

The test was introduced in a very early phase at the first Webnet-Conference in San Francisco and was received enthusiastically by visitors and participants alike. Due to the huge popularity and curiosity of users, the test has already been activated over 1,450 times. This number enables the first analysis of the results. A large number of the visitors of the test site have been European (69.6%), however a large interest has arisen in North America with the Health-Enlightenment (21.3%). Of the 1,450 users, 41.9% were female; 45.7% had an increased risk of Type-2 diabetes. The median age stood at 38.5 +/- 14.4 years. 18.3% have used the test more than once (2-9 times, mean 2.5 times). In a comparison between the Once and the More-than-Once users 25.9% out of 27.8% had a previously diagnosed diabetic condition. 38.3% of 43.8% had a familial predisposition to the disease.

We have shown with the advent of our Risk-Test that the Internet presents an ideal basis for the automation and cost-effectiveness of the publication, dissemination and analysis of risk-questionnaires. Through selective screening and immediate identification of high-risk patients, timely therapeutic intervention is made possible. Thus the Internet offers a new pathway for the education and prevention of Diabetes mellitus.

[1] Herman WH, Smith PJ, Thompson TJ, Engelgau MM and Aubert RE: A new and simple questionnaire to identify people at increased risk for undiagnosed diabetes, Diabetes Care, 18 (1995) 382-387.

UNIVERSITY WEB MANAGEMENT: A DISTRIBUTED MODEL

Phyllis C. Self, Ph.D.
Executive Director for Information Resources & Media
Scherer Hall
923 West Franklin Street
P.O. Box 843059
Richmond, VA 23284-3059
804/828-0634
pself@vcu.edu

In February 1996, the Provost charged the Executive Director for Information Resources & Media to redesign the university web site and to develop web guidelines for the university. Over a four month period three campus wide committees consisting of faculty, administrators, artists and instructional designers developed web guidelines, created a new university-wide web site, and proposed a support system for the VCU Web.

A distributed web management model was established consisting of Information Providers and Technical Contacts from each school. The day to day management of the VCU Website is the responsibility of the VCU Webmaster, Web Coordinator and the Executive Director for Information Resources & Media and a faculty advisory committee. While the web site and the web guidelines are constantly evolving the management has remained the same.

This poster session will highlight the distributed model of web management, the creation of the university-wide guidelines, and the ongoing needs for university-wide support services available to faculty, staff and students. Attendees will be encouraged to review the VCU Web Site at http://www.vcu.edu/web/support/index.html

The Probe Method: A Thorough Investigative Approach to Learning

Glenn Shepherd
Assistant Professor, Educational Technology
Teacher Education
Eastern Michigan University
Ypsilanti, MI USA
gshepherd@online.emich.edu

Poster/Demonstration

The Probe Method is an instructional method that incorporates problem-based learning, interdisciplinary learning, cooperative learning, mastery learning, individualized learning, and the integration of technology with a special emphasis on the use of the Internet. The Probe Method requires students to thoroughly investigate a topic, question, or problem, and in so doing, students learn how to break a complex topic or problem into smaller parts and use the Internet to find the necessary information to understand the topic or solve the problem.

The Internet has provided us with a tool to access enormous amounts of information and to communicate with individuals and experts all over the world. Education needs an instructional method in which the Internet can be most effective. The Probe Method allows students to become fully active in the learning environment. Students learn the steps in solving complex problems. By using the Internet for specific problem-solving tasks, students learn how to learn and how to be critical of what they read on the Internet. In the process, students also learn basic skills, research skills, self-learning skills, problem-solving skills, and communication skills.

The Probe Method was designed by the author and is being implemented in a school system for a dissertation study. This study will collect quantitative data on how the use of the Probe Method might affect critical thinking skills and dispositions toward problem solving. This study will gather qualitative data on how students and teachers feel about the Probe Method and the author will make recommendations for modifications.

The audience for this demonstration will be presented with an overview of the Probe Method and an outline of the steps involved in this approach. They will also receive the most current update to the status of the study and will be asked for their opinions and suggestions for using such an approach on a larger scale.

Directing Student Web Research: No Surfing Allowed

Karen Smith-Gratto, Assistant Professor
North Carolina A&T State University
Curriculum and Instruction: Instructional Technology
Greensboro, NC 27411
gratto@netmcr.com

Gloria Edwards, Assistant Professor
Purdue University
Curriculum and Instruction: Educational Technology
Lafayette, IN 47907
gloedwar@gte.net

With the explosion of the World Wide Web, students have an incredible opportunity to develop information search skills - on-line - in what seems to be one of the largest "libraries" in the world. The idea of exploring a never ending, universal "encyclopedia" which includes untold, and yes, uncensored information is attractive. However, with a combination of classroom time constraints, varying types and qualities of information, and the highly distractive nature of the Web, educators must take a structured approach to teaching students how to use this tool effectively and efficiently. Preparing students to critically sift through mounds of links and information mandates that educators know ahead of time what students might find by (a) clearly defining the purpose of the search, (b) teaching students about the results of broad vs. narrow descriptor searches, and (c) searching the Web themselves prior to sending the class out to explore.

Virtual-U: An Online Learning Environment

Denise Stockley, Chris Groeneboer, Tom Calvert, & Linda Harasim*
Faculty of Education, School of Computing Science, School of Computing Science, & School of Communications
Simon Fraser University, Burnaby, B.C. Canada V5A 1S6
dstockle@sfu.ca , groen@cs.sfu.ca,, tom@cs.sfu.ca,, linda_harasim@sfu.ca
*This work was done collaboratively, and order of authorship is not a reflection of contribution.

Virtual-U is a World Wide Web-based networked learning environment customized for the design, delivery, and enhancement of post-secondary education and industry-based learning. One of the design goals is to provide a flexible framework to support pedagogies based on principles of active learning, collaboration, multiple perspectives, and knowledge building, and to support varied content areas and instructional formats. The framework consists of tools to support core activities including course design, individual and group learning activities, knowledge structuring, class management, and evaluation.

The Virtual-U project is comprised of a multi-disciplinary team of educators, HCI specialists, engineers, computing scientists, database designers, instructional designers, implementors, instructors, learners, and researchers. The system is currently being field tested at 15 universities and industries across Canada to deliver courses from a variety of fields.

Cybersearching for a New Career: Exploring Career Hunting in the Electronic Age

Professor Karen Svenningsen
City University of New York
College of Staten Island
2800 Victory Boulevard
Staten Island, New York 19314-6600
USA
svenningsen@postbox.csi.cuny.edu

Searching for employment in today's fast pace electronic environment can definitely cause chaos for the cybersearcher. Computer services, including the Internet can enhance a job search, but some services available on the World Wide Web are much more useful to job seekers than others. With all the useful enhancements the Net has to offer, within minutes a career hunter can scan the resources of the WWW, traveling abroad in a couple of seconds to seek employment. The searcher can explore an electronic equivalent of classified advertisements, or place a resume into the electronic job market with a couple of keystrokes, quickly and cheaply. One of today's most important resources for the job seeker is a computer with a modem and the ability to distinguish between valuable sites and those not as useful.

Many questions arise with all the possibilities the Net has to offer, such as; Will this be the trend of the future as employment agents become a thing of the past? Does the future hold an extinction for the career placement centers that will not necessarily need to be housed in a physical environment such as a campus.

As faculty applying the Internet resources, this electronic paradise, thrush us to become the explorers to a different world of opportunities. Opportunities for job seekers, researchers, alumnus, fellow colleagues, all exposed to the Internet as an important teaching resource with enormous potential for the future.

Performance of Bursty World Wide Web (WWW) Sources over ABR

Bobby Vandalore, Shivkumar Kalyanaraman, Raj Jain,
Rohit Goyal, Sonia Fahmy
The Ohio State University, Department of CIS
Columbus, OH 43210-1277
Phone: 614-688-4482, Fax: 614-292-2911
Seong-Cheol Kim
Principal Engineer, Network Research Group,
Communication Systems R&D Center
Samsung Electronics Co. Ltd., Chung-Ang Newspaper Bldg.
8-2, Karak-Dong, Songpa-Ku, Seoul, Korea 138-160

We model WWW servers and clients running over an ATM network using the ABR (available bit rate) service. The WWW servers are modeled using a variant of the SPECweb96 [1] benchmark, while the WWW clients are based on a model by Mah [2]. The traffic generated by this application is typically bursty, i.e., it has active and idle periods in transmission. A timeout occurs after given amount of idle period. During idle period the underlying TCP congestion windows remain open until a timeout expires. These open windows may be used to send data in a burst when the application becomes active again. This raises the possibility of large switch queues if the source rates are not controlled by ABR. We study this problem and show that ABR scales well with a large number of bursty TCP sources in the system. The full version of the paper is available at http://www.cis.ohio-state.edu/~jain/papers/webspec.htm

Develop Your Own Multimedia Projects for the WWW

Shelle VanEtten
Research Institute for Assistive and Training Technologies
University of New Mexico
USA
shellec@unm.edu

Tarrae Bertrand-Hines
Research Institute for Assistive and Training Technologies
University of New Mexico
USA
tarrae@unm.edu

Multimedia on the WWW is a powerful vehicle which encourages learners to become active builders of knowledge, rather than just receivers of information. Using multimedia in instruction or accessing information on-line provides learners with new adventures; however, when learners design on-line multimedia, they create and reveal the adventure themselves.

Designing and authoring multimedia can be a powerful instructional strategy for learners. Authoring takes learners beyond reacting to information and allows them to design and organize ideas for others. The design and authoring process requires learners to think deeply and critically about the content they are learning. "Some of best thinking results when students try to represent what they know" (Jonassen, 1996).

Multimedia on WWW includes video, audio, graphics, and text. There are low-cost, user-friendly technologies which are accessible for learners of all ages and abilities that empower learners with access to WWW to actively share ideas that they build.

References

Jonassen, D.A. (1996). *Computers in the classroom: Mindtools for critical thinking*. NJ: Prentice-Hall, Inc.Jonassen, D.A. (1996). *Computers in the classroom: Mindtools for critical thinking*. NJ: Prentice-Hall, Inc. (Jonassen, 1996).

Electronic Teaching and Learning
- New Ways To Learn -
Challenge or Menace for Future Education

Jürgen Vaupel
Manfred Sommer

University of Marburg
Department of Mathematics and Computer Science
Hans-Meerwein-Straße
D-35032 Marburg / Lahn

vaupel@informatik.uni-marburg.de
sommer@informatik.uni-marburg.de

Abstract: Starting with a description of a actual teaching and learning situation at an University this article deals with a possible further evolution of university education. Motivation for this work is the discussion how to integrate new powerful technologies like the Internet and Multimedia in university education. Thus the second part of this article deals with the question what does happen when these technologies are simply applied to an university education system. These considerations lead to a framework for an "electronic lecture". Because of imperative additions to the application scenario, the resulting framework extremely differs from the first one. New components and services like security or on-line examination have to be added to the framework. After implementing this framework for a "virtual lecture" the last step reviewed in this article is exporting and importing educational units to and from other universities or companies. It is pointed out that extending this framework from a "virtual lecture" to a "virtual university" is more a conclusion than a real extension. Adding some centralized services and institutions like payment services and an "education broker" will lead to a third framework, which is an electronic commerce framework for education. This article concludes with some statements about the influence of information technology to university education and the way of learning.

First Framework: The Actual Situation

Since hundreds of years students are sitting in lecture halls, using pen and paper to note what their professors are saying and writing down on the blackboard. In the last few years this situation has changed a little bit while lecturers are using slides and overhead projectors. The lecturer can use prepared slides or even develop new ones "on the fly" and give copies of this slides to the students.

In the meantime an "electronic" version of this scenario is realized at many Universities. This article is based on experiences made at the University of Marburg since several years. Every main lecture ("Informatik I" through "Informatik IV") of the computer science department consists of about 500 electronic slides. Even most of the lectures for senior students are presented with slides. The slides are produced now using Powerpoint (or FrameMaker). To display the slides a lecturer has two principle possibilities. To be flexible in the choice of the lecture hall one can use a NoteBook computer and a portable LCD overhead display. A more advanced (and expensive) method is a special prepared "MultiMedia" lecture hall. In this "MultiMedia" lecture hall a combination of a network PC and a special projection unit is installed. This projection unit copies the contents of the PC display onto an electronic board in front of the audience.

Even this first implementation of an "electronic lecture" changes the preparations of a lecturer in many ways. A lecturer, giving a new lecture, can use the advantages of the digital availability of the slides. Instead of reimplementing the complete new lecture he can use existing slides to modify them and keep them up to date. After the modifications are done, he stores the slides on a network file server, where they can be accessed by the students. At this point of our first implementation of an "electronic lecture" the changes for the students begin. The first obvious change is that they are now able to read the contents of a lecture by electronic means and use all the advantages like searching etc.. They can also print out the lecture and use it as a traditional "old" paper based lecture script, again with all the bundled advantages of this version, like making annotations etc..

Discussing the pros and cons of such an implementation from a lecturers point of view there are two main topics. The first one is the time and work to invest in preparing a *first* version of a new lecture. This initial work is much more comparable with writing a book than with a "classical" preparation of a lecture. Trying to use most of the possibilities of the presentation software, like graphics, illustration or diagrams, results in an enormous work to be done. This leads directly to the second main topic, namely the choice of "tools" to implement such an "electronic" lecture. In the production process today's implementation uses a presentation software to generate the slides and a file server to store and publish them. On the "consumer" side lecturers are using the presentation software loading the slides from the file server and displaying them onto a blackboard. Students are able to read, copy or print the slides. An actual situation at an University can be described with the following application scenario:

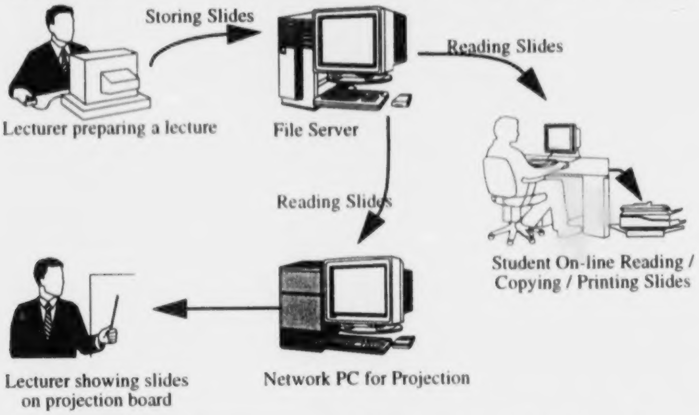

Storing Slides
Reading Slides
Lecturer preparing a lecture
File Server
Reading Slides
Student On-line Reading / Copying / Printing Slides
Lecturer showing slides on projection board
Network PC for Projection

Figure 1: Today's implementation of an "electronic" lecture at an University

Transforming this application scenario into a framework leads to:

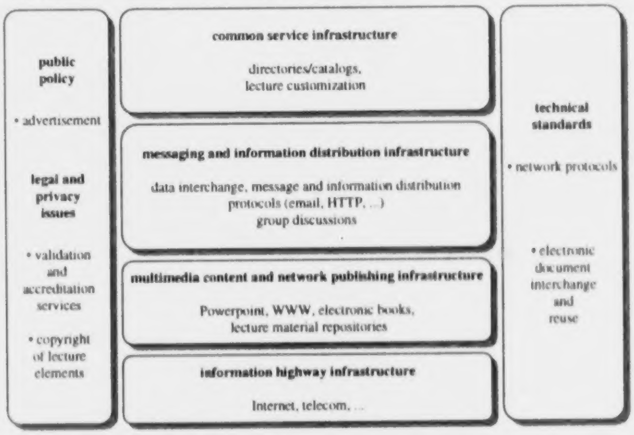

public policy
• advertisement
legal and privacy issues
• validation and accreditation services
• copyright of lecture elements
common service infrastructure
directories/catalogs, lecture customization
messaging and information distribution infrastructure
data interchange, message and information distribution protocols (email, HTTP, ...) group discussions
multimedia content and network publishing infrastructure
Powerpoint, WWW, electronic books, lecture material repositories
information highway infrastructure
Internet, telecom, ...
technical standards
• network protocols
• electronic document interchange and reuse

Figure 2: Actual Framework for a Lecture

This framework consists of three main column. The first one deals with public, private and legal issues, the second concerns different infrastructural subjects and the last one describes the necessary technical standards.

The public, private and legal issues are related to topics, which can be described as organizational. The first topic mentioned in this context is the problem of advertisement. An university or department has to establish mechanisms to ensure that a lecture is properly advertised so that every interested student gets informed. Also a validation and accreditation service has to be provided in order to guarantee that only validated and accredited lecturers can supply lectures to the students. As usual when offering information in digital form, also the general problem of copyright protection has to be taken into consideration.

Reflections on the infrastructure of such a framework are varying in a wide range from a common service down to an information highway infrastructure. Nevertheless the technical possibilities given by a system are mainly defined by the infrastructural circumstances. The choice of the information highway infrastructure determines for example the opportunities of students to access a lecture. The multimedia content and network publishing infrastructure defines the possibilities and limitations for producing a lecture. The messaging and information

BEST COPY AVAILABLE

distribution infrastructure decides on the way the lectures are distributed and published. The common service infrastructure is responsible for a catalog and directory service to support course customizing out of existing lectures. The examples above prove the importance of the infrastructure for availability and quality of service.

The technical standards are of course determined by the infrastructure of this framework and vice versa. One has to choose the proper network protocols in order to support the chosen information highway infrastructure and on the other hand the selected messaging and information distribution infrastructure. The standard for electronic document interchange and reuse highly depends on the multimedia content and network publishing infrastructure as well as on the common service infrastructure. Neither one component nor the other can be changed without paying attention to the crossrelations in this framework.

Nowadays newer and more powerful technologies and services, like the Internet and WWW, are changing the ways business operate and people work. These technological developments are also reshaping the expectations, needs and opportunities in education and learning. The basic information technology tools for developing new ways of education are already available. Especially promising technologies are interactive video, networking and collaboration tools. Access to learning resources has never been as easy as it is via the Internet. Worldwide collaboration is a reality through the World-Wide-Web, creating unprecedented flexibility in time, location, content and form of instruction. But technology alone is not the solution. Reaping the benefits of computers first requires training of the lecturers, new curricular materials and changes of educational paradigms. The experiences with the first application scenario emphasize the need of developing new ways to learn.

Second Framework: "Electronic Lecture"

The emerging technologies that make up the biggest difference in education fall into three broad categories: networking, multimedia, mobility. Integration of these technologies in our first implementation of an "electronic" lecture seems to be easy and straightforward. In a first step we convert the Powerpoint slides into the HTML format, copy these files onto our WWW server and every student around the world is able to attend our lectures. Beside some technical questions, if for example HTML pages are the medium of choice for lecture slides, reasoning a little bit about this approach will raise many fundamental questions. For example:

- What's about the advertisement of these lectures ?
- Who validates the learning material ?
- Are the objectives of the lecture achieved by the students ?
- Is the reading of some electronic slides all that makes a good lecture ?

These questions lead us to the following basic approach for a second framework of an "electronic" lecture in a single university environment:

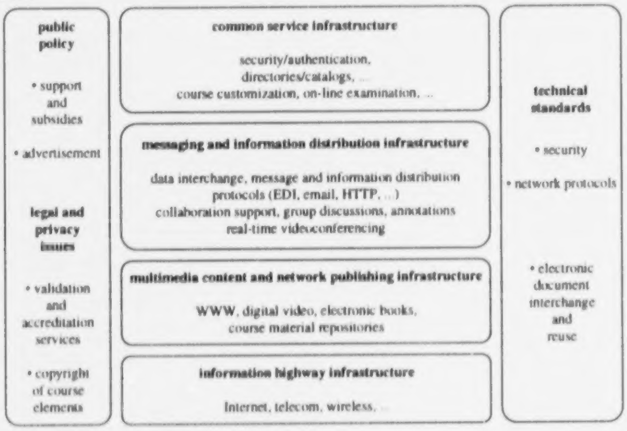

public policy
• support and subsidies
• advertisement
legal and privacy issues
• validation and accreditation services
• copyright of course elements
common service infrastructure
security/authentication, directories/catalogs, ...
course customization, on-line examination, ...
messaging and information distribution infrastructure
data interchange, message and information distribution protocols (EDI, email, HTTP, ...)
collaboration support, group discussions, annotations
real-time videoconferencing
multimedia content and network publishing infrastructure
WWW, digital video, electronic books, course material repositories
information highway infrastructure
Internet, telecom, wireless, ...
technical standards
• security
• network protocols
• electronic document interchange and reuse

Figure 3: Electronic Lecture Framework

It is in the interest of both the lecturers (= producer of education units) and the students (= consumer) to establish standards and quality requirements that apply to the education unit and define procedures for the certification and assessment of the learners progress. A variety of specialized software tools is necessary to ensure customized lectures, as well as tools that facilitate collaborative interaction.

BEST COPY AVAILABLE

Another critical issue is the integration of electronic books with supporting and reference literature. To propel electronic books into a position like the one now occupied by printed learning resources, it is crucial that the ability to find and absorb information be at least as effortless as it currently is with printed learning aids.

Security and authentification mechanisms are also high-priority issues. They need to be adapted to this production process of education units to guarantee the integrity of the learning materials. It has to be ensured that only authorized students have access to the materials and exams and receive credit on completing the requirements.

Accreditation is another major concern in any education-related endeavor. The key criterion is that accreditation must be carried out by an organization endowed by the law with the authority to validate learning material.

Finally, assessment of student progress based on uniform requirements is one of the basic functions offered by all types of education and training. Final assessment of the students' performance is essential for certification purposes.

Setting up LAN's and establishing dial-in services that permit anytime/anywhere access to lecture materials and fellow students eliminates time and space dependencies. New schemes allow students to dial in at their convenience and participate in a lecture asynchronously. While it is not real-time, the opportunity for feedback and participation is enhanced by rich two-way communication channels.

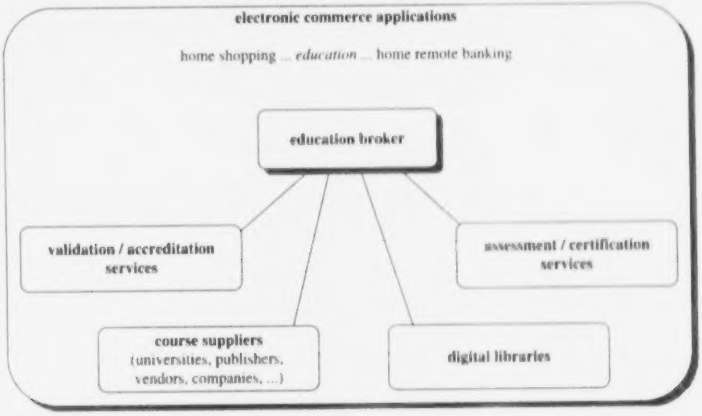
Applying the foregoing ideas will lead us to a scenario which we can call a "virtual lecture". But the ultimate goal is to create a "virtual university" by adding more universities to the application scenario and offer lectures and programs to other locations. This step is essentially more than only increasing the number of involved institutions. Allowing external partners in our second application scenario leads to the requirement of new functionality the so-called "education broker".

Third Framework: "Virtual University - Education Brokerage"

In this new application scenario an "education broker" has to fulfil some of the tasks formerly done by the university like a validation and accreditation service. But the main task for them is to provide product marketing and advertisement for education suppliers. They will match customer needs with existing and prospective education services available from any number of suppliers.

The existing framework for an electronic lecture has also to be extended by an electronic payment system. As education brokerages leverage their ability to mass-market their product to customers around the globe, they will be able to achieve an unprecedented economy of scale that should drastically reduce the unit cost to a mere fraction of what universities commonly charge today. Facing budget pressures universities need to use the technology to reduce costs and increase access to external education suppliers.

With this additions we can review our framework for an electronic lecture as an electronic commerce application. As one of the largest information industries in the world, education has the potential to be a key application in electronic commerce. This leads directly to a framework for a third implementation of an "electronic" lecture, which is an electronic commerce framework for education: ([])


electronic commerce applications
home shopping ... *education* ... home remote banking
education broker
validation / accreditation services
assessment / certification services
course suppliers (universities, publishers, vendors, companies, ...)
digital libraries

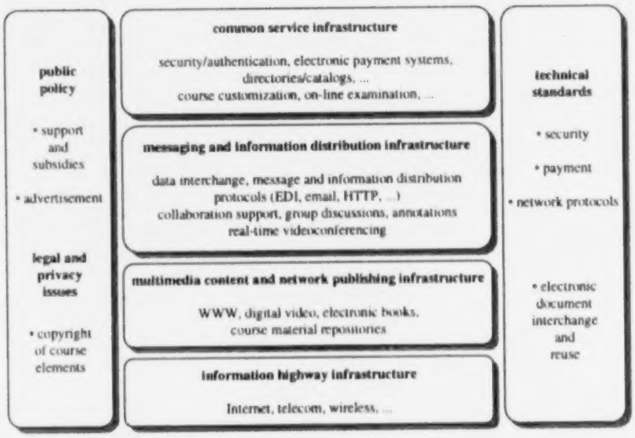

common service infrastructure
security/authentication, electronic payment systems, directories/catalogs, ...
course customization, on-line examination, ...
messaging and information distribution infrastructure
data interchange, message and information distribution protocols (EDI, email, HTTP, ...)
collaboration support, group discussions, annotations
real-time videoconferencing
multimedia content and network publishing infrastructure
WWW, digital video, electronic books,
course material repositories
information highway infrastructure
Internet, telecom, wireless, ...
public policy
• support and subsidies
• advertisement
legal and privacy issues
• copyright of course elements
technical standards
• security
• payment
• network protocols
• electronic document interchange and reuse

Figure 4: Electronic Commerce Framework for Education

All these concepts of distance and distributed learning are applicable to education in industries. This makes "training on demand" possible and brings the information to employees at their workstations. The paradigm of training and learning as a separate activity or centralized department is dead. The new model is learning while working, but new innovative models of production, delivery and presentation are needed that take advantage of the inherent power of this new platform, emphasizing the ability of participants to collaborate globally in real time. The new paradigm is "just-in-time learning" rather than "in-advance learning".

Conclusions

Based on the preceding discussion of the different frameworks and its evolution to an electronic commerce framework for education, a discussion of the conclusions of this work is necessary and will lead us to two main topics: the usability as a classification scheme and the consequences of this evolution for education at an University.

Using the third framework as a classification scheme for multimedia applications and especially for educational applications leads to an efficiency evaluation of these systems. For given systems it can help to find out the pros and cons of a system usage in case of a specific application scenario. During the design of new systems it supports the analysis of the needs and specification of the system requirements in order to produce a complete and satisfactory system. This means on one hand that someone can use the framework to evaluate interesting systems for a given specific application scenario and generate a priority list of abilities which are necessary for his specific scenario. On the other hand it is possible to generate general schemes for systems and produce classes of applications depending on their strength in the different fields of the framework.

The second topic mainly deals with didactic issues in such a way that possible new ways to learn and teach also imply new ways for university education. These new ways are founded in new relationships between teacher and student as well as in new learning paradigms. The representation of education in an electronic commerce framework also establishes new strong relations to other fields with the necessity for education and will raise up many fundamental questions about the quality, sense and contents of university education. Since the development described with the third framework is more a conclusion than a decision, there is a need to answer these questions and depending on this the answer of the question is given, whether the emerging new technologies are a challenge or a menace for future education.

References

[Hämäläinen, Whinston, Vishik 1995] Hämäläinen M., Whinston A.B., Vishik S., *Electronic Markets for Learning*, Communication of the ACM Vol. 39, No. 6, p. 51-58

[Kalakota, Whinston 1996] Kalakota R., Whinston A., *Frontiers of Electronic Commerce*, Addison-Wesley, Reading, Mass., 1996

[Reinhardt 1996] Reinhardt A., *New Ways to Learn*, Byte 20, 3, p. 50-72

[Vaupel, Sommer 1996] Vaupel J., Sommer M., *Electronic Slides and Books*, Proceedings of the 2. Russian-German Symposium: "New Media for Education and Training in Computer Science", Infix 1996

BEST COPY AVAILABLE

Developing an Online Web-based Course

Ronald J. Vetter, Associate Professor
Department of Mathematical Sciences
University of North Carolina at Wilmington
601 South College Road
Wilmington, North Carolina 28403
Email: vetter@cms.uncwil.edu

Abstract: This demonstration/poster presentation describes the design and development of a "totally online" Web-based course. This project is a work-in-progress. It uses the following technologies for courseware development.

- Electronic Mail for instructor/student interaction.
- World Wide Web Pages for distributing multimedia-based course materials.
- Progressive Network's RealVideo software for video clips of instructor lectures.
- WebBoard software conducting ongoing electronic classroom discussions.
- Cornell University's CU-SeeMe software and Microsoft's Netmeeting software for holding virtual office hours via desktop videoconferencing.
- QuizMaker (specialized software that allows one to create interactive tests and quizzes on the Web) to provide instant feedback to students on comphrension of course content.

The course can be viewed online at - http://vetter.cmsfac.uncwil.edu/~vetter/CLASSES/csc475-spr98/

INTERNET TECHNOLOGIES ENHANCE ALLIED HEALTH PROFESSIONALS' KNOWLEDGE

**Dr. Kenneth E. Wright, Athletic Training Education, The University of Alabama,
Tuscaloosa, AL USA kwright@bamaed.ua.edu
Ms. Vivian H. Wright, Instructional Technology, The University of Alabama,
Tuscaloosa, AL USA vwright@bamaed.ua.edu
Mr. Michael Newman, The University of Alabama, Tuscaloosa, AL USA
mnewman@bamaed.ua.edu**

Abstract:
This presentation presents educational technology components, which assist student athletic trainers in studying and preparing for the National Athletic Trainer's Association's Board of Certification (NATABOC) certification examination. This demonstates how The University of Alabama's Athletic Training Education program, in cooperation with the university's Instructional Technology department, implemented off-line browser software to develop an educational curriculum. This curriculum focuses on several skills the student athletic trainer must master before taking the NATABOC certification exam, specifically, the domains of health care administration and professional development/responsibility. A WWW search for viable pages related to the skills was accomplished and using off-line browser software, the selected files were downloaded and organized into a study bank for students to utilize offline. For students and classroom instructors with or without network access, this is a fast, reliable and efficient way to deliver important lessons and related information. The primary advantages of this technological tool are learner flexibility, maximization of content, timeliness and availability, and content retention.

Setting Up a Web Server

For Interactive Engineering Applications

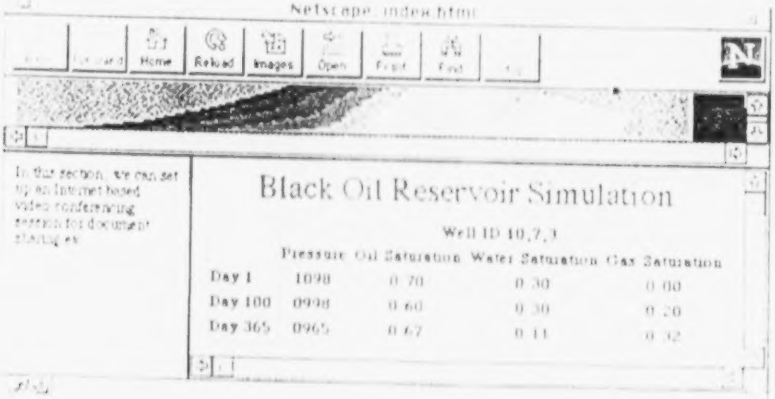
Husam M. Yaghi
Electrical Engineering
Southern University
Baton Rouge, Louisiana 70813 USA
yaghi@engr.subr.edu

John M. Tyler
Computer Science
Louisiana State University
Baton Rouge, Louisiana 70802
tyler@bit.csc.lsu.edu

Poster Session Abstract

We will present a prototype system for utilizing the web for the display and presentation of dynamic results of engineering simulation application. This web server allows us to interactively view results of petroleum reservoir simulations. By making use of CGI scripts, some Java, and Oracle database, simulation engineers can develop a graphical front-end interface to input desired simulation parameters such as well ID, number of layers, number of production days, and output charts. After execution of the simulation program, this advanced web document would use Oracle calls to retrieve the appropriate data, then pass them on to a Java script to animate the results and produce representative graphs.

http://www.subr.edu/~yaghi/FEM



In this section, we can set up an Internet based video conferencing session for document sharing ex

Black Oil Reservoir Simulation

Well ID 10,7,3

	Pressure	Oil Saturation	Water Saturation	Gas Saturation
Day 1	1098	0.70	0.30	0.00
Day 100	0998	0.60	0.30	0.20
Day 365	0965	0.67	0.11	0.32

BEST COPY AVAILABLE

The Internet as a Professional Development and Instructional Resource

College of Education
Idaho State University
Poctello, ID 83209

This paper provides an overview of the applicability of the Internet as an instructional and professional development resource consistent with the National Board for Professional Teaching Standards (NBPTS). These standards frame the broad guidelines necessary to focus effective instructional, pedagogical, and portfolio based activities to insure well prepared and accomplished educators.

The NBPTS board has outlined five propositions that define the "knowledge, skills, dispositions and commitments" that distinguish educators :

1. Teachers are committed to students and their learning.
2. Teachers know the subjects they teach and how to teach those subjects to students.
3. Teachers are responsible for managing and monitoring student learning.
4. Teachers think systematically about their practice and from experience.
5. Teachers are members of learning communities.

Each of these propositions can be directly associated with resources available on the Internet. For example, by accessing selected sites, teachers can find subject matter content, learning theories, instructional styles, and other pedagogically related material such as lesson plans in virtually all content areas. Once downloaded, these resources linked with local resources help a teacher learn how to better to manage and monitor their students' progress. Teachers taught to reflect systematically about their practices and experiences (via portfolios) can also communicate with other practicing or in-service educators through e-mail, open forums, and listserv's. By accessing the resources on the Internet they broaden their learning community from a local one to that of a regional, national, or world learning community.

The Internet is a complex resource on which vast amounts of data and information are stored. Accessing and utilizing this data is an important skill for all educators. Pre-service and in-service teachers are particularly potent utilizers of this resource. The Internet provides a ready resource for educators to frame or supplement virtually any content area and to help develop or strengthen their content knowledge and pedagogical skills consistent with the National Board for Professional Teaching Standards (NBPTS).

In the College of Education, at Idaho State University, teacher education students entering the education program are expected to develop a strong technology orientation. In fact, as they progress through the program they will develop content and pedagogical components that will make up the entries in their electronic portfolios. While numerous entries comprise a given student's portfolio, entries dealing with communication technologies like the Internet and e-mail will be major components of their portfolio.

The use of portfolios as a tool to assess student progress has been a valuable addition to the educational community (Shaklee, 1997). Portfolio assessment allows educators to effectively broaden their formative and summative evaluation processes. A portfolio places in a broader context student performance by allowing the educator to view, often in a nonlinear fashion, the complex set of interrelationships that group to form a more complete picture of student performance. In a similar manner, a portfolio developed by a teacher provides opportunities for that teacher to integrate those complex sets of experiences associated with the development of both content and pedagogical expertise.

It is vital that colleges of education insure that their pre-service training be not only consistent with the NBPTS propositions but give beginning teachers ample opportunity to use the Internet as a dynamic and evolving resource to supplement and extend their understanding and practical attainment of NBPTS standards. Using the Internet as a both a professional development resource and an instructional resource fundamentally addresses the critical interaction between instruction and professional development. And by requiring developing teachers to generate electronic portfolios, this process becomes even more meaningful to the teacher and allows greater integration of their experiences and makes these experiences more authentic.

Presenters:
Bill Yates, Ph.D.
Program Area Leader, Secondary and Technology Education
College of Education
Idaho State University
Pocatello, ID 83209
(206) 236-4353
Yatebill@fs.isu.edu

Stephanie Salzman, Ed.D.
Director, Office of Assessment and Standards
College of Education
Idaho State University
Pocatello, ID 83209
(206) 236-3114
salzstep@isu.edu

Constructing Knowledge in Electronics with the Web

Kai-hing YEUNG
Department of Engineering and Technology Studies
The Hong Kong Institute of Education, Hong Kong
e-mail: lyeung@mc.ied.edu.hk

Research findings indicate that Word Wide Web learning can be facilitated by using a group learning environment. A study was conducted to investigate the effects of group structure on interactions of students working in group with a Web site which contains learning materials on Electronics. Six groups, each of two, three homogeneous ability groups and three heterogeneous ability groups were formed. The talk for each group was audio-taped and transcripts were made to facilitate analysis. It was analyzed and identified as off-task or task-related. The interaction was classified as collaborative work when any of the following was shown: (1) join-task engagement, (2) equality, and (3) mutuality of engagement.

It was found that the learning in group with the Web was effective. Students' collaboration was influenced by the level of difficulty of the task. Also there was a strong evidence of high amount of collaboration when the task is challenging.

IP Packet Filtering Interface Design: Providing Fast and Time Predictable Web Infoshop Services

Chang-Woo Yoon
High Speed Network Access Section
Electronics and Telecommunications Research Institute, Korea
cwyoon@dooly.etri.re.kr

We developed Web Infoshop Service System using the real time OS. This service system has the following features: Open access to Internet, usage based billing, vicarious certification and billing agent service for using charged Web CP(Content Provider).

To provide Web Infoshop Service, we must upload IP packets to TCP, then to HTTP protocol layer. The HTTP is a heavy protocol that requires much processing time compared to IP packet processing. The resource is restricted in the real time applications, so we use packet-filtering concept.

The IAFI(Internet Address Filtering Interface) resides between IP and TCP layer. We formalized the IP packet filtering concepts. We can process only the necessary packets for providing the Web Infoshop service instead of processing all packets. This interface gives high performance and responsiveness of interactive Internet application in the real-time environment.

Acknowledgements: This work is a part of the project, "Web Infoshop Service Node Development", sponsored by Korea Telecom.

Virginia Commonwealth University Events Calendar

James B. Yucha
Web Coordinator
Virginia Commonwealth University
Cabell Library, room 123
901 Park Avenue
Richmond, VA 23284-3008
(804) 828-2192
jyucha@vcu.edu

Virginia Commonwealth University (VCU) sponsors thousands of events each year and prior to 1997 did not have a single source for university events, leaving it to the attendee to search the VCU web site to find their event.

The following features were consider critical when creating the VCU Events Calendar. The Calendar had to be web-based and searchable; present information in a standard format and includes links for additional information; allow for customizable departmental welcome screens; and automate as much as possible.

Today, university events are submitted electronically through a web-based form and consolidated into a single comprehensive events calendar database. Event attendees can now click the "Calendar" button on the VCU Home Page and access all VCU events via an intuitive welcome screen.

The poster session will describe the calendar development process, event submission and review process, system features, planned enhancements, and an overall summary. Also, the poster session will include a demonstration of the events calendar. You are encouraged to visit VCU's Web Site at http://www.vcu.edu/.

Tools for Web Site Management

Dr. Greg Zick
Chairman, Department of Electrical Engineering
University of Washington
Seattle, Washington
zick@mail.engr.washington.edu

Jill Fluvog
Craig Yamashita
Lainey Kahlstrom
IKE Web site management office
University of Washington
ike@engr.washington.edu

The complexity of state-of-the-art Web sites has created a demand for tools to improve the efficiency of site development and maintenance. In the last year many HTML and Web page design and editing tools have become available. If these tools are used effectively, Web site managers, designers, and contributors can focus on gathering and targeting appropriate content for their audiences.

One content issue that arises is management of large collections of multimedia items, such as images and video. We will demonstrate a practical, scalable, high performance tool for the archival of multimedia. "Content" can manage collections of over a million images with response times of less than one second. Web page developers may configure metadata to optimize searching and retrieval of the multimedia objects in their data collection. And… the starter system may be a PC-based (Windows NT) system, so your collection may grow from tens of items to millions…

SHORT PAPERS

Strategies for Effective Use of the Internet and Email for Instruction

Gregory F. Aloia, Ph.D. Professor
Department of Special Education
Illinois State University
Normal, IL 61790, USA
gfaloia@rs6000.cmp.ilstu.edu

Ramesh Chaudhari, Ph.D.
Management Systems Coordinator
Information Systems
Illinois State University
Normal, IL 61790, USA
ramesh@rs6000.cmp.ilstu.edu

Jeffrey P. Bakken, Ph.D.
Assistant Professor
Illinois State University
Normal, IL 61790, USA
jpbakke@rs6000.cmp.ilstu.edu

John Beemsterboer
Graduate Student
Department of Special Education
Illinois State University
Normal, IL 61790, USA
jjbeems@rs6000.cmp.ilstu.edu

In a recent article in USA Today [Marklein 1997] it was reported that 92% of college students have access to computers and that they use email as "effortlessly as picking up a phone. Although the impression from the article is that the college student of today is at the edge of technology, this may not be the case. As reported by Valenza [Valenza 1997] students may appear technologically literate but they are often inefficient and overwhelmed by the technology of the Internet. Her observations are reaffirmed by direct experience over the past few semesters as we have introduced students to both the Internet and email in our courses. Although students may have access to computers and technology, they often use it sparingly, or not at all.

Several strategies have been employed that have helped students to make a smooth and relatively painless transition to the Internet and the use of email. The first step was to provide ample training on the use of the Internet and email. In addition to the regularly scheduled workshops on the "Net" and email offered by the college, additional sessions are scheduled by the instructor and the graduate assistant as well as individual tutoring on a case by case basis. The goal by the end of the second week of class is to have all 200 plus students "connected" to the web and to have successfully sent an email to the professor.

In order to handle over 200 emails, a filter is employed. A filter is placed on the server and directs the student messages to a specific mailbox on the professor's computer. The filter is also programmed to provide an immediate response to each email received. This is a very efficient method of reinforcing the students' first attempt at sending email. The students receive instant feedback and the professor can easily validate that they have completed the assignment. Microsoft Excel is used as a grade book for the course and the email responses from the students are merged with the class rosters for efficient record keeping. To further refine their email skills, students are assigned to small groups (n=7) and given specific topics related to the course content that must be completed via email. Each student must have at least three exchanges that include using attachments forwarding email and establishing mailboxes in the process of completing the assignment.

"Listservs" that focus on the topics in the discipline are provided to the students. Although it is an optional requirement students are encouraged to sign up for one of the groups and participate in the discourse of the discipline. Students are also provided real time group discussions in chat rooms.

Once all students are "comfortable" with the Internet and email, they are provided a session about the use of the library resources. The library staff provides several small workshops. While our library is not yet virtual, students are exposed to specific search engines and strategies for accessing databases and systems external to the campus.

Students are required to use the course's web site for material and assignments. To assist them in using the Internet on a regular basis, specific assignments are put on the Net. For example, bonus questions for the examinations, extra credit assignments, outlines from selected lectures, due dates, class activities, the semester calendar and even a joke of the week are found on the web site. The URL for the web site is http://coe.ilstu.edu/gfaloia

Strategies based on Valenza [Valenza 1997] are presented during the initial training workshops and on the courses web site on how to efficiently "surf the Net" and manage the potential millions of sites that result from inefficient searches. Use of search engines and subject directories as well as Boolean operators, phrases, proximity and nesting are presented. Novelli [Novelli 1997] has identified several ways of saving time on the Internet e.g. skipping the browser page, turning off animation, graphics, and sound which are also discussed.

Ongoing feedback from the students is to solicited insure that the experience is successful and on task. At every class there is an evaluation sheet for students to anonymously ask questions and/or to provide feedback on all aspects of the course. Depending on the specific question an answer is provided via email, the web site, in class, or in person. The primary Internet assignment requires students to identify problems they faced and solutions they employed in the completion of the task. Insights garnered from this feedback are then integrated into the next semester's activities. Students are also encouraged to use email to ask questions and correspond with the instructor.

The feedback form has proven very helpful in providing students with easy and frequent access to the instructor. One of the goals of the class is to provide each student with an "individual" contact with the instructor. The feedback from the students, via email, daily course evaluations and specific assignments has been compiled. The major problems and concerns students face when confronted by the various technologies are found on the web site for the class under the section entitled "Frequently Asked Questions."

In closing, students are also provided with a clear reminder that the Internet and all the related technology are not without their caveats. A quote from Nicholas Negroponte in his work, Being Digital, captures both the caution and optimism of this new, dynamic, and ever changing technology:

> "Bits are not edible; in that sense they cannot stop hunger. Computers are not moral; they cannot resolve complex issues like the rights to life and death. But being digital, nevertheless, does give much cause for optimism. Like a force of nature, the digital age cannot be denied or stopped. It has four very powerful qualities that will result in its ultimate triumph: decentralizing, globalizing, harmonizing, and empowering." [Negroponte 1995, p. 228-229]

References

[Marklien 1997] Marklein, M. B. (1997). At Colleges, email is as easy as ABC. USA Today, March 5, 1997
[Negroponte 1995] Negroponte, N. (1995). Being Digital. Alfred A. Knopf, New York.
[Novelli 1997] Novelli, J. (1997). Speed surfing for educators. Electronic Learning, (16), 5, 52.
[Valenza1997] Valenza, J.K. (1997). Master the art of searching. Electronic Learning, (16), 5, 62-63.

The Effects of Navigation Maps on World Wide Web Usability

Lara Ashmore
Curry School of Education
University of Virginia, USA
lara@virginia.edu

Introduction

The purpose of this study is to investigate the effects of navigation maps for World Wide Web sites on user performance and subjective user satisfaction. Web browsers lack basic navigation support mechanisms such as visual effects to emphasize navigational dimensions that are present in earlier versions of hypertext systems [Nielsen 1995]. The reason navigation is important is because it directly influences a learner's experience with educational materials. If the learner cannot navigate efficiently within the materials, the actual content becomes secondary and jeopardizes the learning experience. Without helpful navigational aides, educational Web users may become frustrated and discouraged, develop negative attitudes towards on-line educational documents and become unable to transfer their browsing experience into knowledge. This study explores the following research questions:

Does use of navigation maps increase user satisfaction?
Does the type of navigation map have an effect on browsing patterns and user satisfaction?
Are Web browsing patterns and user satisfaction associated with certain personal characteristics?
How does task (browsing, searching, looking) affect browsing behavior?

Since the appearance of the first usable hypertext system 50 years ago, there have been volumes of research in the areas of hypermedia systems, user modeling, interface design, human computer interaction, and usability engineering that provide the foundation for present and future research in Web usability. Recent studies on navigation and Web usage indicate that people prefer information to be organized hierarchically [Zimmerman et al. 1996], they tend to revisit few pages frequently [Tauscher & Greenberg, 1996], browse in small clusters of related pages [Tauscher & Greenberg, 1996, Catledge & Pitkow, 1995] and generate short sequences of repeated URL paths [Tauscher & Greenberg, 1996]. People also tend to use the back arrow in browser software 30% of the time and use the bookmark feature less than 2% of the time [Tauscher & Greenberg, 1996, Catledge & Pitkow, 1995] and return back to "entry points" and start over when feeling lost. People have been classified into "search browsers," "general purpose browsers" and "serendipitous browsers" [Catledge & Pitkow, 1995], based on access patterns, but few studies have investigated specific factors associated with Web browser satisfaction to help improve educational materials.

Methodology

This study will employ a randomized field experiment in which self-selected study participants evaluate an educational Web site for a university computer center over a one week time period. The Web site material used for the study is a copy of the Information, Technology and Communication (ITC) Web site at the University of Virginia [ITCWeb 1997] which contains several thousand Web pages about computing information and technical resources. Prior to accessing the experimental Web site, participants will complete a 13 item Web-based pre-browsing questionnaire. The questionnaire asks about past Internet experience, Internet usage, task (browsing, searching or just looking) and descriptive information such as age, gender, and status (faculty, staff or student). Participants will be randomly assigned to one of three treatment groups: no navigation map, a static map or a dynamic map. At the completion of their browsing session, participants will fill out a 25 item Web-based self-administered four point Likert-scale questionnaire. The questionnaire, loosely based on the *Questionnaire for User Satisfaction*, created by interface designers Scheinderman and

Norman [Scheinderman 1992], asks general and specific questions related to user satisfaction and navigation tools. In addition to the survey data, individual performance data will be collected electronically by customized Web server access logs as participants use the Web site. Data will include a participant identifier and traditional usability measures such as the number of pages accessed, time at pages accessed, time spent in help and other browsing actions taken by the user. The identifier allows the participant survey responses to be linked with their actual browsing behavior.

The two navigation site maps tested in this study are hierarchically structured, color, graphic representations of the Web site [Ashmore 1997]. The static map shows only two levels of detail and always remains the same no matter when the user requests the map. The dynamic map [Dynamic Diagrams 1996] generates itself depending on which Web page the user is reading when they request the map. The static site map is a standard image map context-independent representation of the Web site. Characteristics of the static map include: an overview of site; the depth of each section is shown but only top page titles are active and visible. The dynamic site map interface is a *Java* [Sun Microsystems 1996] applet that is a context specific graphical representation of the user's current page location within the site. Characteristics of the dynamic map include: all pages are active and visible; all detail is available; and it shows where the current page is located in relation to entire site.

Results

This is a study in progress. Results will be reported at the conference. Hopefully, the results will help contribute to the growing body of research in the area of Web-based hypermedia and help develop guidelines for user-centered design of Web-based materials and Web browsing tools. In addition, this study will explore new research methodologies and technologies for evaluating Web browsing patterns and user satisfaction.

References

[Ashmore 1997] Ashmore, L. (1997). Example navigation map treatments. [On-line], Available URL: http://amuck.itc.virginia.edu/itcweb/usability/paper/treatments.html.

[Catledge & Pitkow, 1995] Catledge, L. and Pitkow, J. (1995). Characterizing browsing strategies in the World-Wide Web, 1995, *Computer Networks and ISDN Systems*, 27, 1065-1073.

[Dynamic Diagrams 1996] Dynamic Diagrams. (1996). *MAPA*, [Computer software], Dynamic Diagrams, Inc. Providence, RI: Available URL: http://www.dynamicdiagrams.com/.

[ITCWeb 1997] ITCWeb (1997). [On-line], Available URL: http://www.itc.virginia.edu/.

[Nielsen 1995] Nielsen, J. (1995). *Multimedia and hypertext: the Internet and beyond*. Boston, MA: AP Professional.

[Schneiderman 1992] Schneiderman. (1992). *Designing the user interface*. New York: Addison-Wesley Publishing Co.

[Sun Microsystems 1996] Sun Microsystems. (1996). *Java*, [Computer language], Mountain View, CA: Sun Microsystems, Inc.

[Tauscher & Greenberg, 1997] Tauscher, L. and Greenberg, S. (1997). Revisitation patterns in World Wide Web navigation. *Human Factors in Computing Systems*, 1997, ACM Press, Atlanta, GA. In press.

[Zimmerman et al. 1996] Zimmerman, D., Muraski, M, Palmquist, M., Estes, E., McClintoch, & Bilsing, L. (1996). Examining World Wide Web designs-lessons from pilot studies. *Designing for the Web: Empirical Studies*, 1996, Microsoft, Redmond, WA. [On-line], Available URL: http://microsoft.com/usability/webconf/zimmerman.htm.

Acknowledgments

This research was made possible by the encouragement, generosity and support of Dynamic Diagrams, Inc. and the Office of Information Technologies at the University of Virginia.

Creating Web Presence using NetPresence System

Aaron Aw, Chan Fang Khoon, Jimmy Tan, Louisa Chan, Tan Tin Wee
Internet Research and Development Unit,
National University of Singapore,
10 Kent Ridge Crescent
Singapore 119260
E-mail: aaron@irdu.nus.sg

Introduction

The invent of the World Wide Web (WWW) has created a whole new chapter in the history of the Internet. This WWW architecture (HTTP, HTML, Hyperlinks, Browsers etc) which supports multimedia information has revolutionized the way people interface with the Internet. However, due to the inherent stateless characteristic [Mower 1996] of the protocol, it is difficult to provide dual-way communication between users on the Internet. There is no interaction between the webpage owner and the websurfer who is browsing it except through the use of electronic mail. A simple way of creating a communication channel through the web can be implemented based on a "NetPresence Model" proposed by our unit.

NetPresence System

The NetPresence System (NPS) is designed to allow Internet users to find out who is on-line [Tan and Chan 1997]. It also allows an Internet user to broadcast his presence in the Net once he is on-line. We have implemented a prototype NPS using a client/server model. This prototype allows a user to detect or be notified of someone's presence on the net provided he is using the NetPresence system. If a person is found to be on-line, a short message can be sent to him/her. Figure 1.0 shows the model of our NetPresence System. We proposed a protocol, NetPresence Protocol [2] to be adopted in all Internet applications. With the protocol incorporated, you can register with the NetPresence (NP) server when you start an Internet application and announce your presence to a list of users/friends. You can then "look up" on any user registered with the NP server to obtain the location of the user and the application that he/she is running.

For example, a user running a "NP enabled" Internet Relay Chat (IRC Client) application will register his/her presence to the NP server by using a command "/np_announce" and the IRC client will send an "Announce" message to the NP Server. This message will register the user's presence by providing user information such as USER ID and IP Address. Once his presence has been registered, the server will update the user's "Friends" list so that he knows which of his friends are online. The user will perform a "/np_exit" to remove their own presence when they stop using the Internet. To facilitate a fast and convenient communication channel among the NP users, there is also a messaging feature which allows short messages to be sent from one user to another, or to several users simultaneously.

The NP protocol can even be further improved such that the NP server can still reach you by some contactable means, even when you are not logged on. For instance, you can instruct the NP server to page you if a particular user that you are looking for has just logged on to the net. Also, you may want to leave a message with the NP server, so that this message can be directed to anyone who is looking for you on the net, even when you are not logged on.

Web Presence

From the previous section, the NetPresence concept can be implemented on various Internet applications such as IRC, Telnet, MUD etc. In addition, we feel that the current WWW interface can also be incorporated with this concept. One example of using this Web Presence is in a technical/system support department of a

distributed organization. It can be used to locate the support staff and messages can be sent immediately to the staff instead of using E-mail.
The Web Presence System will make use of the NetPresence server using a Java-based applet client component. Figure 2.0 shows the implementation. We are able to support a variety of platforms due to the Java nature of the client system. The Web Presence client serves as the WWW interface to the NetPresence system. This would allow any WWW user to contact or find out who is currently logged on to the NP system. The WP client would require less download time and this is particularly useful for one-time communication with the other party. As the client makes network connections to the NetPresence system in the back-end, the web client can communicate with other applications that support the NetPresence protocol. An example of this is WWW-IRC communication.

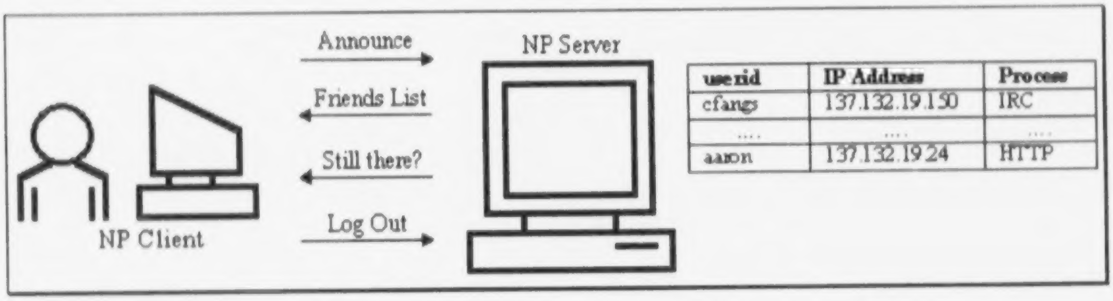
The execution of the Java applet would allow the Web Presence system to provide a real-time chat/messaging function, complementing the short messaging function in the previous NetPresence System. The Web Presence client would also include a "Friends" list where the user would be able to know which of his friends are on-line. Using the Web Presence system, a registered user is also able to make use of the facilities provided by the NetPresence system. For example, the NetPresence system can act as a file transfer conduit for two Web Presence clients. This would facilitate file transfers between the two clients. User authentication is also built into the Web Presence system whereby a user is required to register/authenticate himself before he can use the client. A Log-In client would perform the log-in process before the actual applet client is activated. The registration process would involve the user providing his User Identification (UID) and password. This information is sent encrypted through the network to the server. The NetPresence server would then verify this information with the user database. Once his identity is verified, the actual Web Presence applet client would be sent to the user.

Conclusion

The implementation of the NetPresence System has been exciting. We have created a simple but yet important prototype to address the current issue of lacking communication on the WWW. Future developments of the system will include enhancing the server component for seamless communication to other applications (e.g. WWW-IRC, IRC-WWW). Further enhancements are also being made to the client to provide more functionality.

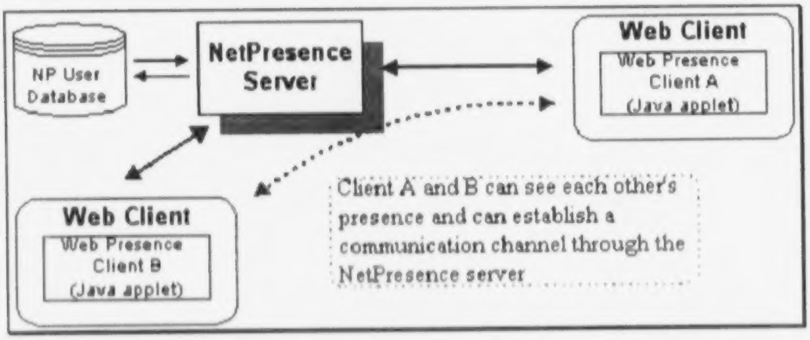
References

[Mower 1996] Mover, Matt (1996). Stateful and Stateless service, http://axis.unl.ac.uk/prjdoc/node31.html
[Tan and Chan 1997] Tan,Jimmy & Chan, Louisa. NetPresence Protocol, http://irdu.nus.sg/npp



userid	IP Address	Process
cfangs	137.132.19.150	IRC
....	....	....
aaron	137.132.19.24	HTTP

Figure 1: NetPresence System model


NP User Database
NetPresence Server
Web Client
Web Presence Client A (Java applet)
Web Client
Web Presence Client B (Java applet)

Client A and B can see each other's presence and can establish a communication channel through the NetPresence server

Figure 2: Web Presence implementation of the NetPresence system

Development of Simple Campus Intranet Courseware of English as a Foreign Language for Japanese Learners

Junichi Azuma
Faculty of Commerce
University of Marketing and Distribution Sciences, Japan
PAG02676@niftyserve.or.jp

Kazuhiro Nomura
Faculty of Information Sciences
University of Marketing and Distribution Sciences, Japan
HGG03562@niftyserve.or.jp

Introduction: How WWW is Used in English Classes in Japan

People throughout the world witnessed the "Internet Explosion" which was triggered by the invention of the WWW in the mid-1990s. The new type of network communication by the WWW has been rapidly spreading throughout the world and the word "Internet" seems to have come to denote "WWW" itself these days. In Japan, too, the popularity of the WWW and the easy access to the Internet by dial-up connection have tremendously increased the number of the Internet users these days.

Owing to the emergence of the WWW, we can enjoy seamless access to various multimedia resources of all over the world, as if we were browsing the pages stored in our own computer in the hyper-text format. The potential of the WWW inevitably attracted educators all over the world, since it has the power to integrate educational resources scattered around the world into something like a "tailor- made" multimedia database created for their own educational purposes.

Teachers of English as a foreign language who are teaching in non- English-speaking countries have eagerly welcomed the development of the WWW and easy access to the Internet, since they can obtain a vast amount of the timely teaching resources written in English just by a click of a mouse. In Japan, where English is widely taught as a foreign language, more and more enthusiastic teachers of English are trying to introduce into their classes the material obtained through the WWW. A lot of Japanese universities or colleges have recently established the campus LAN system which ensures the full Internet access. These days special classrooms for English with one networked computer per student are not rare in these institutions.

In this situation, however, something very unnatural seems to be happening in Japan. Sometimes, teachers tend to rely on the materials obtained through the Internet too much, forgetting to develop their own teaching materials and making light of the classroom environment where the face-to-face communication with the students is really possible. It is all right to show the students the materials obtained via the Internet. It is also important to teach the students how to navigate in the vast ocean of the Web world. But what has become of the actual role of the teachers? Is it really sufficient just to show the students the list of URLs where the learning materials exist? Is it enough to let students learn English via the Internet in the "high-tech" classroom while the real person-to-person communication is possible in the same room? Furthermore, how can we cope with the "traffic jam" of the Internet which is supposedly caused by so many people accessing the popular sites?

Proposal of Development of Campus Intranet Courseware

One of the ways to resolve the above-mentioned unnaturalness in the teaching environment and to lessen the "traffic jam" of the Internet is to develop the teaching materials in the HTML format by ourselves and to store them in the HTTP servers in the institution so that the students will have the easy access to the material from the classroom or other places where client computers are available. In addition, these materials will be also accessible from at home by way of the dial-up connection services.

Teachers can create their original HTML format teaching materials mainly as assignments to be done outside class. During the regular classes they can then concentrate on the face-to-face exercise activities.

The transfer rate of this kind of the Intranet system is considered to be quite high, since the students are not expected to navigate outside of the campus LAN into the Internet world, unless some links to the outside sites are offered in the materials created by the teachers.

How to Develop Simple HTML-Based Teaching Materials

The development of the HTML-based teaching materials for the campus Intranet system is not so difficult as one might imagine. If the teachers would like to develop really interactive courseware, complicated techniques will be necessary, including the cgi programming. However, if we are satisfied with simple "pseudo- interactive" materials, then we have only to learn several specific techniques.

For example, to practice rapid reading, one may want to control the timing of the text presentation, i.e., showing a page of five or six lines for 10 seconds, then presenting the next page automatically. This type of the rapid reading materials can be easily realized by using the tag <META HTTP-EQUIV="Refresh" CONTENT="X; URL=FILENAME.html">, where X shows the time (seconds) before the next page is loaded, and "FILENAME.html" indicates the file name (with the appropriate path) or the URL of the next page.

To present the result of the marking of the answers onto client machines, teachers can create the marking procedure by JavaScript. Forms for the answers of a multiple-choice type exercise can be prepared beforehand so that the marking mechanism may be activated by a certain event handler.

One of the shortcomings of JavaScript programming is that if students are mature users of the Web browsers, they can easily find the source of the HTML document including the part where JavaScript is embedded. In such a case, they can guess the answers of the exercise by detecting the part where the correct answers are shown. One of the best ways to avoid this type of "cheating" is to use Shockwave for the marking routine.

In the actual presentation, the authors would like to provide some examples of rapid reading exercises developed by JavaScript and Shockwave locally in order to show that these types of the HTML-based courseware can be used not only in the TCP/IP environment but also in the different network environments such as AppleTalk or Windows95 network, and even in the stand-alone environment.

Conclusion

Especially in Japan, comfortable "Net-Surfing" has become very difficult because of the nation-wide "traffic jam" of the Internet. However, the authors consider that now is the best time for teachers to conduct researches on the development of the HTML-based campus Intranet courseware. What will be gained by these researches can be easily transformed into the "Internet" courseware when the real "Information Super Highway" is available.

Virtual Workshop: Interactive Web-Based Education

Kathy Barbieri, kab23@cornell.edu
Cornell Information Technologies, Cornell University, Ithaca, NY, USA

Susan Mehringer [1], susan@tc.cornell.edu
Cornell Theory Center, Cornell University, Ithaca, NY, USA

Introduction

Over the past ten years, the Cornell Theory Center (CTC) [2] has earned national recognition for developing and delivering high performance computing (HPC) education to its national research community. Recently, CTC has enhanced its online workshop materials, and developed a series of Virtual Workshops (VWs) [3] which are offered entirely over the Web. This Web-based approach reaches a larger audience, leverages staff effort and also poses challenges for developing new and meaningful presentation techniques. We could not meet the national demand for HPC education through our on-site workshops. HPC is a rapidly changing topic and requires continual update of materials. In order to maintain leading-edge education, we needed to dynamically update materials (even as they were being offered) and the Web provides this solution. The creation of the VW has provided an asynchronous learning environment which addresses the above challenges. CTC has offered a series of VWs to more than 500 researchers.

Interactive Web-based Features

Self-Referencing Glossary

HPC education introduces many new terms, which are often poorly defined or not defined at all. We have created a self-referencing glossary [4] (using JavaScript) which explains all new terms in the modules. These terms are presented in bold, italicized font. Selecting one of these terms results in a new glossary window popping up on the screen. The selected term and definition are at the top of the window.

Feedback through Quizzes

Most workshop participants want the opportunity to self-test their understanding of the material at frequent intervals. In addition to lab exercises, we have used CGI scripts to write interactive quizzes [5]. The quizzes are written as forms in a multiple choice format. Filling out and submitting the quiz form automatically grades the quiz and returns the results. An option button allows the participant to choose whether they wish to receive a detailed explanation of all quiz answers along with the grading results. Workshop participants like the simple format and immediate answers, as well as being able to test their understanding. In a recent VW, workshop participants rated quizzes very highly.

Presentation Layers

The instruction modules are designed to have two layers of detail [6]. The top layer, called the presentation layer, covers the material in a brief manner appropriate for a speaker to use during a presentation, or for a reader to use as review. The second layer is a detailed or discussion layer. One can choose to read either layer or to move between layers via links provided on each page. Two-thirds of VW participants report that they only use the discussion layer. The remaining third use both the presentation and discussion layer.

Discussion Forums

Face to face workshops provide a natural forum for the participants to ask the speaker questions and for them to discuss common areas among themselves. In an effort to promote this interactive nature, we introduce an email alias 'vw-consult'. We encourage participants to send in their questions via email and we provided answers within 24 hours. The quick turnaround was critical in minimizing their frustration. In an effort to increase interaction among the participants and with CTC staff we introduced a VWMOO and VWChat to a recent VW. We provided general consulting in the VWMOO from 8-5. We offered scheduled office hours for specific topics in both the VWMOO and the VWChat. We also experimented with moderated module readings of two modules. Results were a bit disappointing. Those who were accustomed to MOOs participated freely. Others found it cumbersome to learn a new discussion forum. The VWChat proved more intuitive to use and flowed better with the Web-based VW. However, the scheduled consulting times were contrary to the asynchronous nature of the VW and the times were inconvenient for some. Most important, participants preferred to concentrate on the base materials, and felt those materials and email consulting adequately served their needs. We are pursuing other collaborative forums for future use in the VW.

New Features Under Assessment

Web-based Editing and Program Submission

We have developed a Web-based interface [7] which allows one to edit a program, compile, submit the program and have the results return to the Web page. This easy to use interface keeps them on the same familiar Web page while experimenting with program changes. This interface was developed in conjunction with Northeast Parallel Architecture Center (NPAC) in Syracuse, NY.

Flexible Approach for Different Learning Styles

Use of frames allows workshop participants to move through material in a way that best suits their learning style and needs. In this module [8] we display the Table of Contents in the left-most frame, the main text in the largest frame, and a program in the third frame. This module was designed around an HPF program; workshop participants can learn the material either by working through the program or by working through the topics as displayed in the Table of Contents. Frames also allow coordination of material shown in two frames, by use of a simple link. For example the reader can read about a program directive in one frame, and with a link, force the corresponding material to be displayed in the second window.

Enhancing Learning Through Animation

Some concepts are more easily understood with an animation. We have created an animation [9] of shifting an array (using gifmerge) which enhances the text explanation. A QuickTime movie [10] was created to compare the performance speedup of a parallel program using three techniques. The movie zooms in on interesting portions of the graph as the author is explaining them. As an alternative to the movie, we created a series of three graphs with the corresponding text next to them. Based on limited feedback, the animations did succeed in adding to the value of the explanatory text. The QuickTime movie, while excellent quality, was too large to download over most internet connections.

Conclusions

While we were able to leverage our efforts by using existing training materials and online documents as a starting point, extensive additional effort was required to bring detailed content and interactivity into the VW. We found that quizzes, fast consulting response, and lab exercises were critical, popular, pieces to providing feedback and interactivity. The VWMOO and VWChat did not fare as well, being somewhat redundant, unfamiliar components. VW participants have reported that they felt they learned about the same amount from the VW as they would have from an on-site workshop, and somewhat more from the VW than from an introductory textbook with exercises. Overall VW evaluation ratings have been very gratifying.

URL References in this Paper

[1] http://www.tc.cornell.edu/~susan
[2] http://www.tc.cornell.edu/
[3] http://www.tc.cornell.edu/Edu/VW/
[4] http://www.tc.cornell.edu/Papers/mehringer.mar97/glossary.gif
[5] http://www.tc.cornell.edu/Edu/Talks/HPF/Intro/distribute.html
[6] http://www.tc.cornell.edu/Edu/VW/Talks/MPI/Pt2pt.II
[7] http://www.tc.cornell.edu/Papers/mehringer.mar97/hpfweb.authb.gif
[8] http://www.tc.cornell.edu/Edu/Talks/HPF/Intro/frames.html
[9] http://www.tc.cornell.edu/Edu/Talks/HPF/Intro/cshift1.html
[10] http://www.tc.cornell.edu/Edu/Talks/HPF/Intro/timing.html#Donna

References

Roslyn Leibensperger
Summer '96 Virtual Workshop: Topics in High Performance Computing
Internal report, Cornell Theory Center

Acknowledgment

We would like to gratefully acknowledge our colleagues who have helped develop the VW, especially Anne Trefethen, Associate Director Computational Support, and Ros Leibensperger, VW Project Leader. In addition we thank those involved in the development of the features described above: Donna Bergmark, Doug Elias, Caroline Hecht, Brian Lewis, Chris Pelkie, Nikki Reynolds, and Catherine Starkey.

We appreciate the collaborative efforts of Northeast Parallel Architecture Center (NPAC) in developing the web-based submission particularly Geoffrey Fox and Kivanc Dincer. We thank Cherri Pancake, Oregon State University, who provided expert advice on the usability of the HPF module.

PSYCHOLOGICAL ANALYSIS OF SPECIFIC DEVELOPMENT PROBLEM "USERS FROM FORMER SOVIET UNION IN INTERNET AND WWW SPACES"

Valentina M. Bondarovskaia, Ph.D.
Institute of Applied Informatics, Department of Psychology and Ergonomics, Kyiv, Ukraine
bond @rozrada.freenet.kiev.ua

Introduction. The Internet and World Wide Web culture is a new possibility for Humanity to find the common language, to perceive the different cultures and to surmount many contradictions. All nations and cultures can represent theirselves in these fantastic new artificial worlds. There are some new psychological and ethic problems. Users of Internet can have a free access not only to the achivements of culture and education, career on-line home page but even secret codes of Pentagon (as Zagreb teenagers, for example). In these artificial worlds there are the dramatic clash of different cultures, different ethic systems etc. It is important to study psychological and ethic models of posttotalitarian personality and the influence of psychological and ethic models of postsoviet people for their behavior in Internet.

Objective. The objective of our investigation was to study the psychological and ethic issues of personality inclination to the totalitarian social space, to carry out a research of postsoviet human models that influencing person's attitudes to different cultures, to global ethic norms, human rights, property etc. We analysed the psychological characteristics of postsoviet computer users of different age, their attitudes to legislative problems of using software and the main human ethic norms that are important in new information technologies global worlds.

Methods. We used methods of cross-cultural and psychohistorical analysis for apportionment of psychological and ethic models of posttotalitarian persons. We also carried out the cross-cultural psycho-ergonomical analysis of human-computer interfaces that were created in the developed countries and used in post-soviet infonation space.

The main direction of analysis were: cognitive, emotional, ethic, pattern-proper(the system of attitudes and values).

On the cognitive level of the analysis we considered the peculiarities of people perceptive activities in different cultures, the developing of the strategic thinking, decision making and planning procedures in the different countries and their relationships with modern computer culture using. We analysed the speed characteristics of computer users in the post-soviet information space (reading speed, the information processing speed) and compared these data with speed requirements in the new information technologies modern levels.

We also analysed the soviet and postsoviet personality information space and its psychological and ethic influence for the persons of different age entry to the Internet world. We paid special attention on the postsoviet teenagers entry to this world.

We established the method battery for people psychological state monitoring: 1) the Kosugo's test for measurement of people psychological status with 8 parameters: anxiety, depression, general and chronic fatigue, physical break-down, irribality reflecting conditions of loading, weakened vitality, failing moral; 2) Spilberger's method of the anxiety level appraisement; 3)the method of stress level appraisement -"The scale of situational fear and emotion"; 4)The projective test "Unfinished phrases"; 5) questioning of subjects about their attitudes.

Subjects: 120 computer users and 100 teenagers.

Results. We picked out the psychological and ethic models of post-soviet computer users on the all levels of analysis. We desighned the psychological and ethic picture of computer user in post-communist society. Inclination of personality to the totalitarian social space had the such common psychological ussues as double moral, not distinguishing oneself from class, declining the responsibility about own life, unpretentiousness, absence of responsibility for the property of other persons and organizations etc.

Law for the ordinary citizen didn't exist as something necessary to know, understand, follow or expect to defend his or her rights. Justice during the "Soviet" period was only a secret, punitive mechanism that often killed innocent people. People learned to evade the hostillity of the law and to substitute in their mind a hope for individuals with a conscience of honesty and decency.

One of the dominating psychological constructs of Soviet people was fear as the main inside manager of human behavior. "Fear" combined with "Lie" created one of the most powerful and enduring consequences of totalitarism. Not only did fear force people to suppress the sincerity in relationships between friends, it eventually paralysed the development of sociological, psychological, political and ecological sciences.

Patternalism was also one form of the main features that has the ordinary postsoviet person. In the whole on the cognitive level of modeling we can see that cognitive space of soviet personality had such characteristics: limited information space, the absence of the freedoom, the fear as the constant emotional life manager, the lie as the way of the constant society information space functioning, the lie as the way of the person surviving, unmoulding of the personality strategic thinking, unmoulding of the planning skills, unmoulding of "the time feeling" etc. All these characteristics have the direct influence on the person's behavior in WWW space. It is time to establish the preventive psychological strategies for formation safe entry of postsoviet persons of different age to Internet and WWW space.

We measured the emotional status of computer users that work in computer space the different time - from 4 hours per day to 12 hours per day. We picked out the ergonomical peculiarities of the modern human-computer interface that have the specific emotional issues to emotional status of users, especially teenagers. The agression level is higher not only under the influence of the content of the computer game, for example, but under the influence of the image color or its dynamic's characteristics.

As the preventive strategy we propose to share in the new information technology space the special psychological and ergonomics knowledge and technologies that will help many people to reduce the aggression and anxiety level not only of their own products but also will raise the culture of using the new information technologies in their families.

THE INTERACTIVE LEARNING CONNECTION - UNIVERSITY SPACE NETWORK (ILC-USN). A HYBRID INTERNET/CD-ROM COURSE IN SPACECRAFT SYSTEMS DESIGN.

AUTHOR: Dr. William Brimley, P. Eng.
Co-ordinator, Media in Education,
Adjunct Professor, School of Aerospace Engineering
Rogers Communications Centre
Ryerson Polytechnic University,
350 Victoria Street,
Toronto, Ontario, Canada
M5B 2K3
E-Mail: bbrimley@acs.ryerson.ca
Website: http://www.ilc-usn.kcc.ca

The Interactive Learning Connection - University Space Network (ILC-USN) is a successful implementation of technology enabled distance learning. A consortium of Canadian Universities (including University of Windsor, University of Western Ontario, York University, Ryerson Polytechnic University, Queen's University, and Royal Military College), partnered with Ontario Centres of Excellence (which include the Institute for Space and Terrestrial Science/Centre for Research in Earth and Space Technology, Knowledge Connection Corporation, Information Technology Research Centre), industry (Spar Aerospace Limited), and a Resource Centre (Marc Garneau Collegiate Institute - Space Resource Centre) to launch a "Spacecraft Systems Design" pilot project.

The pilot project initially consisted of multimedia learning modules on CD ROMs, each module focusing on a particular topic of spacecraft systems design. Each module was authored by a subject matter expert from a participating university. The course has quickly moved onto the Internet, with all modules now converted to HTML, and available through the ILC-USN website <http://www.ilc-usn.kcc.ca> .

Modules/Authors include:
- SPACECRAFT SYSTEMS - W. Brimley - Ryerson
- ORBITAL MECHANICS - P. Somers, J. de Boer - RMC
- SATELLITES & PROBES - P. Somers, T. Racey - RMC
- PROPULSION SYSTEMS - R. Sellens, P. Oosthuizen, J. Bryant - Queen's
- MECHANICAL - W. Brimley - Ryerson
- ROBOTICS - R. Buchal - Western
- ROBOTICS ASSEMBLY AND MAINTENANCE - L. Reeves, A. Hopkinson - RMC
- ELECTRICAL - W. Brimley - Ryerson
- SPACE SOFTWARE - L. Reeves, A. Hopkinson - RMC
- GROUND CONTROL - J. Soltis - Windsor
- DESIGN OF RELIABLE SYSTEMS - H. Jack - ex Ryerson

ILC-USN is now using a hybrid application of HTML authored modules available both on the web (with password protection) and CD-ROM. Full video and audio is best accomplished (with present user bandwidth restrictions on the Internet) by the student browsing a local CD, and alternately browsing the web for the latest module up-dates, and links to other suggested sites.

The implementation of the pilot project, and evaluations by third party reviewers, resulted in refinements to the course. These refinements have been implemented in the present Development Phase of the project, and include the following; An ILC-USN server and website, Internet conferencing, use of e-mail, FTP, student websites to submit assignments, student final reports mounted on websites, threaded newsgroups, and evaluation forms.

Student response and enthusiasm for the project is excellent. Well over 100 Engineering and Science students (fourth year and graduate) have taken the course in the four offerings since the fall of 1995. Students at each university comprise a Team, and collaborate in conceptualizing and designing their own spacecraft which must meet demanding functions specified at the start of each course. For many students, this is the first chance they get to work as a Team. At each site there is a site coordinator (usually a professor who has authored one of the modules) to provide the human-to-human student/professor interaction. The students enjoy the Team Learning approach with its rich environment for development of personal interaction skills, and demand to all share the same mark at the end of the course. Their final mark is based on marked assignments plus a final report which integrates their corrected assignments. At the completion of each course the Teams from all universities gather together at one of the sites, to meet each other (often for the first time), and to present their Final Report to the other Teams. These gatherings are exciting for USN students and staff, and a rewarding experience for all participants. Teams also provide their Final Report on the web.

We firmly believe that the ILC-USN has become a model for Technology Enabled Learning and Collaborative Team Learning. The ILC-USN is now expanding to include more universities across Canada, the United States, and Mexico. Courses being added include French Language modules, a Space Policy and Law Course (in French and English), an Engineering Graphics Course, and a Remote Sensing Course.

Universities expressing interest for future collaboration in this next phase (Phase B) of the USN Project include;

- York University (Jan-April 1998)
- Royal Military College (Sept-Dec 1997, Jan-April 1998, Sept-Dec 1998)
- Queen's University (Sept-Dec 1997, Sept-Dec 1998)
- Ryerson Polytechnic University (Sept-April 1997/98, Sept-April 1998/99)
- Ecole Polytechnique (Sept-Dec 1998 English/French)
- Simon Fraser University (Sep-Dec 1997)
- NADI (North American Design Institute) (Sept-Dec 1997, Sept-Dec 1998) to provide students from Universities in the United States (Detroit Mercy, Santa Clara), Universities in Mexico (Monterey Tech, Guadalajara), and Canada (Simon Fraser, Ryerson).

KEY REFERENCES

1. [ISTS 1994]. Summary Report of the USN Founders' Conference.
2. [Brimley and Rowe, 1994]. W. Brimley and I. Rowe (1994). The University Space Network. *4th Annual Canadian Student Space Conference, 8th CASI Conference on Astronautics, Ottawa, Nov 7, 1994.*
3. [Brimley and Rowe, 1995]. W. Brimley and I. Rowe (1995). A New Learning Paradigm - The Interactive Learning Connection. *Networking for Profit ISTS Showcase, September 20, 1995.*
4. [Gillis and Pike, 1996]. L. Gillis, R. Pike (1996). Innovations in Course Delivery for a Distributed University - the Evaluation of the ILC-University Space Network Pilot Project, *KCC Report.*
5. *ILC/USN Conference Proceedings*, May 6, 1996.
6. [Brimley and Rowe, 1996]. W. Brimley and I. Rowe (1996). The University Space Network, *Education and Space Conference, 9th CASI Conference on Astronautics, Ottawa, November 13, 1996.*
7. [Brimley, 1997]. W. Brimley (1997). The Interactive Learning Connection - University Space Network (ILC-USN). A Hybrid Internet/CD-ROM Course in Spacecraft Systems Design, NAUWeb'97, June 1997.

Distance Learning - Courseware Development

Netiva Caftori
Computer Science and Women Studies
Northeastern Illinois University
Chicago, IL., USA
n-caftori@neiu.edu

As population increases and the need for education is made more evident in this changing world the demand for distance and web-based courses will only grow [Bork, 1997]. Major learning modes in schools and universities are still the lecture and the textbook. Since people are different and require different teaching styles we know too well now that these major learning modes are outdated.

Computers provide us with new opportunities for learning with their capability of interactivity and connectivity. The web in particular opens wide doors for obtaining information and seeing things with new lenses. However we need to be cautioned as information is not knowledge and knowledge acquisition alone is not learning [Rudenstine, 1996]. What we want learning to be is the utilization of knowledge acquired in the solving of problems.

A good example for us to follow of teaching via problem solving is found in the Netherlands. Courses at the University of Maastricht are conducted in the style of Problem-Based-Learning (PBL) which is derived from the Harvard University model [Caftori and VanReeken, 1995]. Classes are restricted to 12 students. The instructor acts as a program and project manager, but does not lecture. The students conduct the classes, taking turns as the class secretary and summarize the activities of each class in written form, which is then presented at the next class. Students identify the cases and problems they need to explore, then report back as to how they achieved their results.

This PBL technique is important to the study of on-line education, as the lecture mode is not desirable on-line either. Rather, students can be directed to a group of problems and cases, depending on the subject, which is summarized in the syllabus and course guide, written by the instructor. Students then work in groups to achieve results, and report back to the class in written form.

Another model for us to follow as a guideline to good teaching are the seven principles of Good Practice [Chickering and Ehrmann, 1997]. Good Practice encourages contact between students and faculty, reciprocity and cooperation among students, active learning techniques, prompt feedback, time on task, high expectations, and respect for diverse talents and ways of learning. We contend that all these principles can be observed using the new distance learning technologies combined with committed faculty and students who are made aware of them.

Too many web pages available for teaching today are designed in the lecture mode of presentation. We would like to caution our audience about this danger. We would like to encourage individualized teaching as much as possible by including interactivity whenever available and presenting problem situations and group work.

Web courses provide 24-hour access, increased interaction between students and faculty, and more flexibility in learning styles. Face-to-face meetings are still essential. When students meet in the classroom, notes need not be taken since students know that all materials are on-line already. There is therefore less passivity on their part and more interaction and active learning. Web courses can provide good assessment of students' progress as well by installing counters which count the number of visits to certain sites and quizzes which give immediate feedback. Keeping student records and progress is important for individual learning and for assessment of progress and its monitoring. Privacy can be maintained on-line as well although the technology is not user-friendly yet.

Some of the benefits of web courses include the sense of community that students gain, the increased attention that they exhibit in class because of the opportunity they are offered to prepare ahead of time for class, the flexibility of space and time to learn, and the possibility of cross-platforms. Since interactions on-line may now be based on common interests and not just on physical space, one surprising new benefit is that students spend more time studying on topics of concern to them and therefore more learning is taking place.

Another important factor in reaching students is the natural language we use in traditional classrooms. Natural language is our most powerful tool for communication. It lends itself to complex learning. Whenever possible we should include this mode of communication in our distance learning as well. Conferencing on-line, listservs, e-mail, chat environments, news-groups, telephoning, "moos" and "muds", or face-to-face meetings are examples of such human interactions. It is important to use some of them in some form. Use voice recognition devices whenever available.

Smart classrooms, as found at Northeastern Illinois University, combine traditional classroom setting with networked computers. The sky is the limit as to what can be achieved in such an environment: From hands-on learning, to team work, to individualized attention to class discussions and presentations.

This is still a brainstorming era. Collaboration is essential among instructors since many hurdles are on our way. Copying source code of HTML, Java or CGI from each other is one way of collaboration. Preparing a web course is time consuming. Most courses available present therefore only a syllabus. We depend on the technology and the support we receive is usually insufficient. Helping each other can have far-reaching effects.

As a summary, distance learning can be conducted in many different ways: by correspondence, video conferences, or web-based courses. It allows people with handicaps, people who live far away, people with similar interests, or people with different life-styles to obtain an education at a distance. It is up to a good teacher with much preparatory work to conduct teaching and learning in a very effective way. By using modern technology but keeping basic pedagogical methodology, while implementing the "Seven Principles for Good Practice in Undergraduate Education" [Chickering and Ehrmann, 1997] one can achieve what we consider quality education.

References

[Bork, 1997]. Bork, A: (1997). The Future of Computers and Learning, T.H.E. Journal, 25th anniversary issue

[Caftori and VanReeken, 1995]. Caftori, N. and VanReeken, A. (1995). Problem-Based Learning, Information Systems Education Conference, 1995, Data processing management Association, Charlotte, NC.

[Chickering and Ehrmann, 1997] Chickering, W. and Ehrmann, S. (1997). Implementing the Seven Principles. Technology as Lever, http://www.aahe.org/ehrmann.htm

[Rudenstine, 1996]. Rudenstine, N.L. (May 29, 1996): Special Address, Harvard University.

Making the MOST of Virtual Reality

Margaret D. Corbit, Cornell Theory Center, Cornell Univ., USA, corbitm@tc.cornell.edu

Introduction

Virtual reality (interactive, 3D, computer-generated) environments promise to add a new dimension to Web communication, holding great potential for enhanced interactivity and exploration. With the advent of the Pentium processor for the personal computer, viewers today can navigate in 3D spaces on the desktop. One of our goals in our outreach efforts at the Cornell Theory Center (CTC), the NSF supercomputing center housed at Cornell University, is to take advantage of this advance in technology to present scientific content in an engaging way based on VRML.

Working in a Controlled Environment: a Museum-based Exhibit

Several institutions developed on-line exhibits for the Museum of Science and Technology (MOST) in Syracuse, New York, as one aspect of a highspeed networking collaboration among the supercomputing center and school of education at Syracuse University, Rome Laboratories, NYNEX (the regional communication provider), CTC, and the museum. The project included installation of a computer lab on the floor of the MOST. In coordination with educators at Syracuse, the project includes training the museum staff and docents as well as developing methods for evaluating the sites. CTC's contribution took advantage of this high-bandwith computer lab to devise a compromise testbed for experimenting with VRML.

The combination of a high-speed connection and a "captive" audience (groups of children and adults visiting the museum and monitored by museum staff) offered us the opportunity to ignore software, platform, and networking constraints and to focus on the medium and the content. In addition, we plan to experiment with what is being referred to as hybrid CD technology, for example Netscape's LiveCache, which will allow us to put large files on a CD ROM that will sit in the viewers machine, and alleviate download lags. We hope this will lead to developing sites that can be viewed at libraries and remote science centers.

Mapping the Gaps at MOST (http://www.tc.cornell.edu/er96/MOST)

NYGAP, part of the national Gap Analysis Program (GAP) of the National Biological Services, was featured in CTC's 1995 online science book in an article that included VRML files translated to relatively manageable sizes as illustrations. Because NYGAP is a New York State program, it provided an appropriate focus for our exhibit at MOST. We conducted further background research with the help of Smith and NYGAP researchers which allowed us to extend and enhance the feature as a 3D exhibit.

The exhibit is analogous to an extension of the gallery space in the museum, a room with similar floor and wall coverings and poster-sized images on the wall. Each image is a link to an external browser window that presents the related content, such as information on the bird diversity of the region, including images and animated clips. The calls of local birds provided by the Cornell Laboratory of Ornithology are incorporated into the VRML space as ambient 3D sound, luring the viewer to continue exploring. Content is organized by aspects of the project and presented on five of six walls in the room. These walls feature digital technology, geographic information systems, conservation biology, and the collaborators in the program (external links to pertinent pages). In addition, there is a virtual desktop computer station in the room from which browsers can enter CTC's online science book.

The sixth wall presents a large image that links to an additional VRML file, a vegetation and land use model based on Landsat imagery. CTC visualization producer Chris Pelkie worked with NYGAP researchers to overlay information from the GIS database on the 3D topography of the state. The file used in the MOST exhibit is a small section of the original model, representing the Finger Lakes region and including the Syracuse area as well as Ithaca, CTC's home base. The viewer can fly over the 3D landscape, seeing the pattern of forest, field, and water in the region between the two cities. Inevitably, everyone tries to find some familiar landmark.

Current Status

We began training staff and volunteers in July, 1997, introducing more than 20 people to the site and it's technology during our first visit. The exhibit space is now open for public access during museum hours when a volunteer is available to staff the room. Our installation originally required use of Pentium PCs running Window 95. The site can now be viewed using Netscape Communicator on either Mac or PC. According to MOST Director of Education, Rachel Nettleton, the *Mapping the Gaps* is the most popular exhibit in the lab.. Initial discussions of monitoring/evaluation methods are under way. Graduate students in Computer Science at Syracuse will be responsible for updating and maintaining the networking. Volunteer highschool students are maintaining the browser.

Note: Browser technology moved more quickly than we anticipated, and we were forced to upgrade the world to VRML2 format in August. At that time, we also began managing a sister site for remote support of the museum staff and volunteers. NYNEX provided financial support for this effort.

Putting 3D Scientific Files on the Web

Getting useful 2D illustrations is always a challenge for science communicators, even when the technology is trivial. Adding another dimension to the files we use is probably an order of magnitude harder. Because most research files are too complex to mount on the Web, we face a sociological problem, in addition to the technical challenges. We have to find researchers willing to invest precious time and energy creating reduced files with lower resolution or reduced extent (i.e., focusing on a small region of the data) in coordination with our capabilities. For this project, we were lucky; CTC staff created the original research files and were available to rework them. Computer graphics students and visualization specialists are devising creative ways to make this process easier at the same time that they are developing a new technology.

Long-term Goals

VR is most often used as an exploratory tool in research, often yielding insight into the nature of the system being studied. When a researcher explores a familiar file with you, they explain concepts by wandering around in the file until they find examples to help. We believe that it should be possible to share the researcher's experience of VR on the Web by creating custom files derived from the original data, and by incorporating text and sound files explaining not only the content but also the way in which the file was generated and the researcher's exploration of it. We have a long-term goal of developing a site around a 3D file instead of tagging the 3D file to less interactive content. When this paper was first submitted for review, we believed that we could not accomplish this goal in the near future; now it looks as though we may be able to mount our first attempt during 1998. The geometries of the files created by researchers using our resources are entirely too complex to render quickly on a desktop, so one focus of our work will be to simplify the files and/or their presentation. In the mean time, we are focusing on getting the enhancements working.

Acknowledgments

National Science Foundation, NIH, NY State, IBM, Corporate Partners, NYNEX. Development Team: D. Cane, T. Chi

FishNet: Finding and Maintaining Information on the Net

Paul De Bra* **and Pim Lemmens**
Department of Mathematics and Computing Science
Eindhoven University of Technology
PO BOX 513, 5600 MB Eindhoven
The Netherlands
debra@win.tue.nl, wsinpim@win.tue.nl

* Paul De Bra also affiliated with the "Centrum voor Wiskunde en Informatica", Amsterdam, and with the University of Antwerp, Belgium.

Abstract

This short paper presents a tool for keeping a hotlist or homepage up to date. It combines two existing tools:

- MOMspider [Fielding 1994] is a tool to verify whether links are still valid and whether documents they point to have been modified or moved.
- Fish-Search [De Bra & Post 1994b] is a search tool for finding new interesting documents in the neighborhood of a given set of (addresses of) documents.

FishNet keeps track of the evolution of a domain of interest by periodically running MOMspider and FishSearch and presenting the user with newly found documents. The user can put documents in the hotlist or in a reject list. This positive and negative feedback are constantly used to improve the precision of the search.

1. Overview and Motivation

Beginning World Wide Web users start collecting addresses of interesting documents they find, by storing them in the browser's bookmark list. Later they may also move this information to their home page to share their findings with the world. Keeping the list consistent and adding addresses of new interesting documents to ensure that the list remains a valuable resource can quickly become a full-time job.

Existing large search engines such as Alta Vista, Excite or Lycos, do not offer a solution to this problem, because no small set of keywords is sufficiently discriminating to perform a search without returning a high rate of non-relevant documents. Browsing through the answers of these engines, in search of some new interesting documents, often takes much more time than it's worth.

The **FishNet** toolkit offers a platform for automating hotlist maintenance. It offers the following features:

- Configuration and Maintenance through HTML forms and Java applets.

- Verification of link consistency and of updates to documents through the standard MOMspider package [Fielding 1994] (developed by Roy Fielding, not by us).
- Multi-threaded Fish-Search navigation engine [De Bra & Post 1994a, De Bra & Post 1994b] for finding new documents. This engine can be extended by means of external filters for determining relevance of documents.
- A set of filters for finding related documents.
- History of documents previously marked as relevant or non-relevant, to improve the selection of new documents.
- (HTML) Report generator through which the bookmark list or home page can be updated.

By means of FishNet the user can ensure that the list or home page always contains valid links, that the descriptions of these documents remain accurate, and that new documents on the topics of interest are found and added to the list. Using FishNet can reduce the full-time information discovery job to just a few minutes a day.

For use with FishNet the Fish-Search tool has been improved significantly since its original development back in 1994. The most important new features of Fish-Search are:

- Fish-Search used to be integrated into a Web-browser. The new version is a stand-alone program that can be activated as a CGI-script.
- Use of multi-threading (through the standard W3C library) to avoid long delays when a slow site is encountered during a search.
- Fish-Search now obeys the Robot-Exclusion protocol [Koster 1994].
- External filters can be used in addition to the built-in keyword- regular-expression and approximate maching algorithms. These filters must reside in a special directory on the server on which fish-search is activated, this to avoid abuse.

Figure 1 shows the global architecture of FishNet and how it fits into the Web and Internet environment.

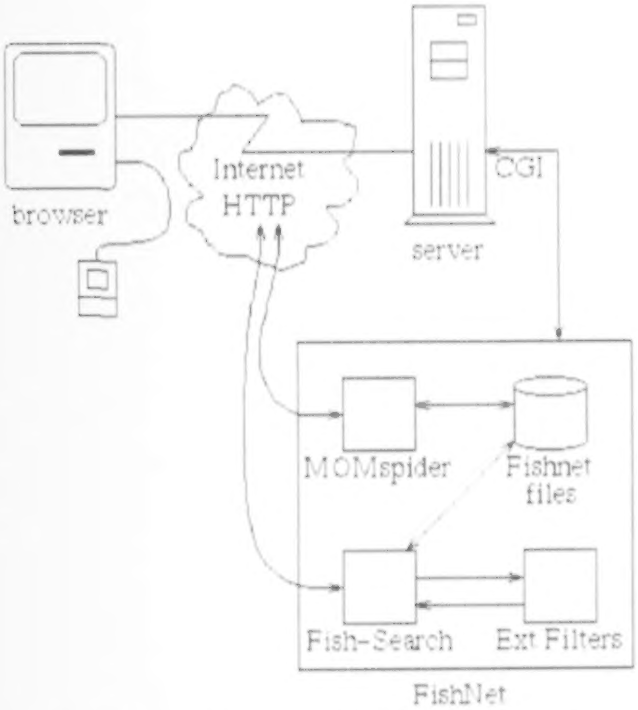


Figure 1: Global architecture of FishNet and its environment.

---

2. Using FishNet

FishNet is normally run at night, from the Unix *cron* utility. It first activates MOMspider to find which documents need closer examination. It then performs the following actions:

- For documents that have been relocated, FishNet updates the hotlist to note the new address.
- Documents that have been modified are starting points for a search-run, in order to look for new interesting documents. FishNet comes with a set of filters for finding related documents.
- For documents that have been deleted, or possibly moved without leaving a relocation, FishNet will start a search from the root of the server(s) these documents used to be on. If the documents were simply moved chances are they will be found again.
- New potentially interesting (URLs of) documents are combined into a report for the user. From the report the user can move the documents to the hotlist or to a reject list.

If FishNet is run through a proxy cache [De Bra & Post 1994a] and the user's browser goes through the same cache, the documents that need to be examined by the user can be retrieved very efficiently.

Some systems try to locate information based on a user profile [Balabanovic et al. 1995]. Some systems even try to deduce the user profile from the browsing behaviour [Brown & Benford 1996]. Since a user may be interested in more than one subject, it is more difficult to determine which information satisfies the user profile than when only one specific topic is used. Some packages like those described in [Maarek & Shaul 1995] and [Gaines & Shaw 1995] try to distribute documents over a set of topics automatically. FishNet does not deal with multiple areas of interest. Instead, for different subjects separate lists or Web pages should be created, and each of the lists is treated separately by FishNet. In order to do so, FishNet identifies each "job" by the user identification and the URL of the list.

The filter package that comes with FishNet is still under development. It currently offers the following features:

- It can determine the language a document is written in.
- It can decide how closely two documents resemble each other by comparing word usage.
- It can generate a "vocabulary" from a set of documents to be used for finding new documents with a similar vocabulary.

By creating a vocabulary for the documents in the bookmark list, and another vocabulary for the reject list, documents can be ranked by similarity to the "good" vocabulary and dissimilarity to the "bad" one.

We believe FishNet is a valuable tool for teaching students about hotlist maintenance. For mainstream end-users some commercial maintenance and search tools are entering the market, with more user-friendly interfaces.

3. References:

[Brown & Benford 1996]

Chris Brown, Steve Benford, Tracking WWW Users: Experience from the Design of HyperVis, WebNet'96, World Conference of the Web Society, pp. 57-63, San Francisco, 1996.
http://aace.virginia.edu/aace/conf/webnet/html/174.htm

[De Bra & Post 1994a]

P. De Bra, R. Post, Information Retrieval in the World-Wide Web: Making Client-based searching feasible, First International World Wide Web Conference, Geneva, 1994.
http://www.win.tue.nl/win/cs/is/reinpost/www94/www94.html

[De Bra & Post 1994b]

P. De Bra, R. Post, Searching for Arbitrary Information in the WWW: the Fish-Search for Mosaic, (Poster and demo at) Second International World Wide Web Conference, Chicago, 1994.

http://www.win.tue.nl/win/cs/is/debra/wwwf94/article.html

[Fielding 1994]

Roy T. Fielding, Maintaining Distributed Hypertext Infostructures: Welcome to MOMspider's Web, First International World Wide Web Conference, Geneva, 1994. http://www.ics.uci.edu/pub/websoft/MOMspider/WWW94/paper.html

[Gaines & Shaw 1995]

Brian R. Gaines, Mildred L.G. Shaw, WebMap, Concept Mapping on the Web, Fourth International World Wide Web Conference, Boston, 1995. http://ksi.cpsc.ucalgary.ca/articles/WWW/WWW4WM/

[Koster 1994]

Martijn Koster, A Standard for Robot Exclusion, (Unofficial standard obeyed by most robots on the Web). http://info.webcrawler.com/mak/projects/robots/norobots.html

[Balabanovic et al. 1995]

Marko Balabanovic, Yoav Shoham, Yeogirl Yun, An Adaptive Agent for Automated Web Browsing, Stanford University Library project, Working Paper SIDL-WP-1995-0023. http://www-diglib.stanford.edu/cgi-bin/WP/get/SIDL-WP-1995-0023

[Maarek & Shaul 1995]

Y.S. Maarek, I.S. Ben Shaul, Automatically Organizing Bookmarks per Contents, Fifth International World Wide Web Conference, Paris, 1995, Computer Networks and ISDN Systems, Vol. 28, p. 1321-1335. http://www.ics.forth.gr/~telemed/www5/www185/overview.htm

Task-centered Navigation in a Web-accessible Data Space

Eckehard Doerry, Arthur E. Kirkpatrick, Sarah Douglas, Monte Westerfield*
University of Oregon
Dept. of Computer Science, *Institute of Neuroscience
Eugene, OR 97403

Within the World Wide Web (WWW) community, navigation aids are usually designed to make it easier to find information of interest. Powerful search engines have been developed to winnow through millions of pages in search of potentially relevant information. Formal models of navigation, such as [Furnas 97], are used to suggest network structures which minimize the number of steps required to reach a goal.

These efforts assume that the users know what information they need but don't know where to find it. The metaphor is one of locating a destination within a data space. While this is a common situation, it is not the only one. We are developing a Web-accessible database for biologists studying zebrafish development [Westerfield et al. 1997]. In this context, the users know exactly what kinds of data are available and where it is located. Our users have few problems moving to information of interest, but they often have difficulty *remembering how they got there*. This paper describes an approach to orienting the users within the *task space* rather than the data space.

The navigation difficulties arise from several characteristics of the biologists' tasks:

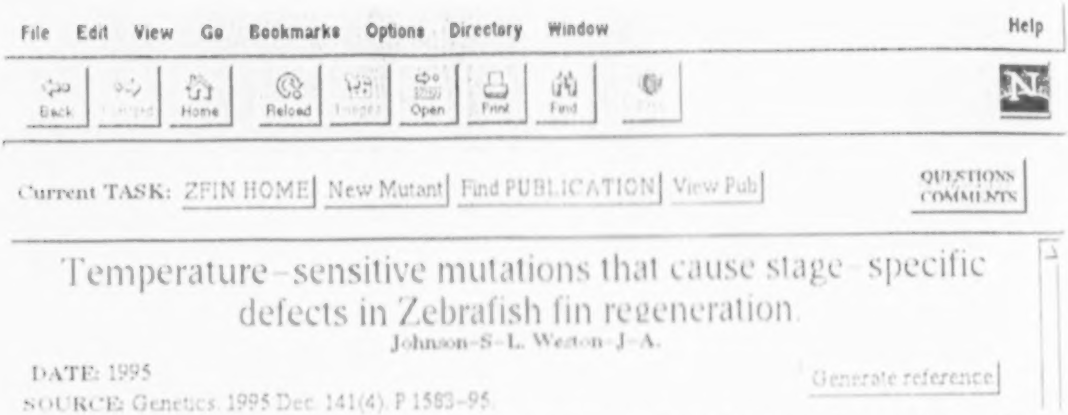
1. **Complex, multi-step activities.** Our database supports direct submission and updating of experimental data by the researchers themselves; this entails an open-ended sequence of nested form-filling, browsing, and selection tasks. For example, to submit a new mutation to the database, a user must specify the lineage of the mutation, the lab at which it was discovered, the mutant's observable characteristics and chromosomal abnormalities, and the publications in which it has been described. Thus, movement through the data space involves intertwined sequences of searching and browsing, rather than a simple unidirectional progression from the data space entry point to target data. The complexity is in the *sequence* of subtasks.

2. **Prevalence of similar displays.** While screens vary in (potentially important) details, they are often similar in overall appearance. For example, every screen that presents a search interface has similar provisions for specifying search criteria and displaying search results. While interface consistency makes the system more easily learned, we have found it to be a confounding factor for navigation.

In this data space, we have found the biologists have little trouble determining what data and activities are available, but easily become disoriented once engaged in an activity. Specifically, they often become confused about where they are within a multi-step process, how their current activity relates to an overall goal, and how to return to previous steps in the process.

Our first navigational aid was a visual representation of the standard browser history list of traversed pages; we found this approach to be completely inadequate because it was not presented in terms of the user's domain level task. For example, biologists searching for a specific class of mutations will iteratively refine their search criteria, making several queries. This leads to a lengthy sequence of pages, all related to the same overall domain task. Conversely, similar searches may occur at different times within an interaction, each associated with a different domain task. In both cases, a simple history list does not reflect the conceptual structure of the domain level tasks the user was engaged in

Motivated by these observations, we are exploring a task-centered model of navigation. A task-centered navigational aid represents the user's current position in terms of the task/subtask hierarchy. We describe the



Temperature–sensitive mutations that cause stage–specific defects in Zebrafish fin regeneration.
Johnson-S-L. Weston-J-A.

DATE: 1995
SOURCE: Genetics. 1995 Dec. 141(4). P 1583–95.

Figure 1: A (truncated) snapshot of a user submitting a new mutation to the database. The user is currently engaged in the subtask of searching for and specifying the primary publication in which the new mutation was described. The task-centered navigational aid appears at the top of the page and reflects the user's position within the task/subtask hierarchy.

users' location in terms of the sequence of their goals rather than the sequence of pages they have traversed. We have implemented a prototype of such a navigational aid for our zebrafish database. The aid appears as a sequence of tiles that reflects the user's current position within the task space. In [Fig. 1], the user is submitting a new mutation to the database. The navigational aid, displayed in a dedicated frame at the top of the page, lists the sequence of subtasks the user has followed. From the left to right, the user started at the home page, began submitting a new mutation, searched for the publication announcing the discovery of that mutation, and is currently viewing that publication. Upon selecting an article to associate with the submission, the user will be automatically be returned to the "new mutant" submission form (the pending super-task), with the selected publication filled in as the "primary publication". The user may also click on any tile in the task path to cancel some current subtask and return directly to a previous step.

Our usability tests of this feature are very encouraging: most users immediately abandon the browser's mechanisms, the "back" button and history list, in favor of the task-centered tool. In addition, users report a much clearer idea of how each subtask fits into the overall task, and how to back up if they change their mind.

While the task-centered approach appears promising, some difficulties remain. A particularly challenging problem is how to gracefully accommodate arbitrary digressions. In the course of entering new data, for example, a user may notice that an existing record is incomplete and digress from the current task to update that record. The navigational aid must somehow recognize and display such digressions in its representation of the user's position within the task space.

The increasing popularity of complex, web-accessible data spaces demands navigation aids more powerful than history lists and ubiquitous "return to home page" buttons. In particular, users may need assistance orienting themselves within the task steps. For domains with a well-defined task space, we believe that a task-centered model of navigation can provide an effective framework for maintaining that orientation.

References

[Furnas 97] Furnas, G. W. (1997). Effective View Navigation. *Proceedings of CHI '97*, ACM, New York. 367-374.

[Westerfield et al. 1997] Westerfield, M., Doerry, E., Kirkpatrick, A. E., Driever, W., & Douglas, S. A. (1997). An on-line database for zebrafish development and genetics research. *Sem. Devel. Biol.* (in press)

Acknowledgments

BEST COPY AVAILABLE

The zebrafish database project is supported by the W. M. Keck Foundation and NSF grant BIR-9507401.

Integrating Security Services Into Collaborative Systems

B. S. Doherty
Department of Computer Science & Applied Mathematics,
Aston University, Birmingham B4 7ET, United Kingdom.
(b.s.doherty@aston.ac.uk)

M. A. Maarof
Department of Computer Science & Applied Mathematics,
Aston University, Birmingham B4 7ET, United Kingdom.
(maarofma@aston.ac.uk)

Abstract: Computer Supported Collaborative Working (CSCW) provide computer support that facilitates co-operation between users. The increased attention on CSCW brings with it a need for security in the development of group applications. Much work has been done on the technological aspects of CSCW [Foley and Jacob, 1995] but the aspect of information security of CSCW technology has not received much attention [Teufal et al. 1995], which encourages research on the security issues in CSCW. The emergence and wide-spread adoption of WWW offers a great deal of potential for the developers of collaborative technologies, both as an enabling infrastructure and a platform for integration with existing end-user environments [Bentley et al. 1995].

In this paper we are proposing the incorporation of security services into CSCW system operating over the Internet. The main purpose of the security services is to allow groups of users to securely access the CSCW system, and also to secure information flow between users participating in the system. The security services provided will cover only the synchronous distributed mode environment (same time but at different places). Distributed Code Inspection Groupware [Doherty and Sahibuddin, 1997] will be used as a case study for CSCW application in implementing the security services.

The security services that will be incorporated into CSCW system are *group access* and *encryption* services. Group access services provided will allow groups of users to securely access the CSCW working systems. A secret sharing protocol or threshold scheme [Shamir 1979] will be used in implementing group access to the CSCW system. In this protocol a key (k) is divided into n shares ($k_1, k_2, \ldots, k_n$). Knowledge of any quorum q or more shares allows k to be easily computed, while knowledge of $q-1$ or less shares leaves k completely undetermined. Hence, this protocol will not only provide secrecy and reliability, but also safety and convenience[Shamir 1979]. This protocol can be implemented in variety of setting for CSCW application.

In a co-operative situation, it is desirable that communication between members of the group is secure [Sakakibara et al. 1994]. Encryption services will be incorporated into the CSCW system to ensure security of the information flow between users participating in the system. Public-key cryptosystems is not suitable for real-time processing (synchronous interaction) since public-key cryptography involves computation between very large integers and is highly CPU intensive [Freirer and Karlton, 1996]. A combination of public-key and symmetric cryptography will be used in implementing encryption services. This combination provides flexibility of public-key cryptography with the speed of symmetric cryptography [Koblitz 1987]. In this technique, keys can be exchanged using the slower public key cryptography; while the large volume of messages would be sent by the faster, symmetric cryptography. Data Encryption Standard (DES) will be used for message encryption in the encryption service. DES are commonly accepted standard common encryption scheme, well known and well established [Karila 1991]. Kerchoff's principle states that only published encryption algorithms have potential security, because they have been investigated and exploited thoroughly.

The two services will be provided as a layer between web layer and CSCW application [Fig. 1]. This layer act as a gateway for accessing the CSCW application. By including these security services as a layer between the application page and the web layer, application will be secured. On accessing the CSCW system by the group synchronously, users will be authenticated. At he same time, key from the validated users will be computed before granting access to the CSCW application. Quorum of shares can computed the key and granted access to the system. Less than quorum of shares specified leaves the system unaccessable. The encryption services will ensure security of the information flow between users participating in the system. By incorporating the security services into the CSCW system, transactions and information flow between members of the group will be secured. This security services will be implemented into Distributed Code Inspection Groupware [Doherty and Sahibuddin, 1997].

Current Distributed Code Inspection Groupware (DCIG) under development [Doherty and Sahibuddin, 1997] extend the current technology of code inspection groupware, but the system does not consider securing the code inspection groupware system or its information: the source codes. In their system, the proposed security system works where the members of the code inspection process access the system synchronously. By using the WWW as the platform, which provides easy flow of data anywhere in the network, the information (source codes) will flow freely but without any protection. The lack of security would limit the application of the system proposed by Doherty and Sahibuddin. By integrating the security services into DCIG application will enhanced it to be secured. The proposed security system will be developed using Java and is intended for all system that use the WWW to communicate. By using security mechanisms provided by Java, the complication of developing the prototype of a secure application on the WWW can be tackled.

As a conclusion, we have presented two security services to be integrated into the CSCW System which will enhance the security of CSCW systems. Someone who is using the system will be confident that the system is secured from being access by unauthorised users and that information that flows on the network will not be tapped. As for Distributed Code Inspection Groupware this security services will enhance the current technology of code inspection groupware by making it secure, especially the source codes which is the most valuable asset.

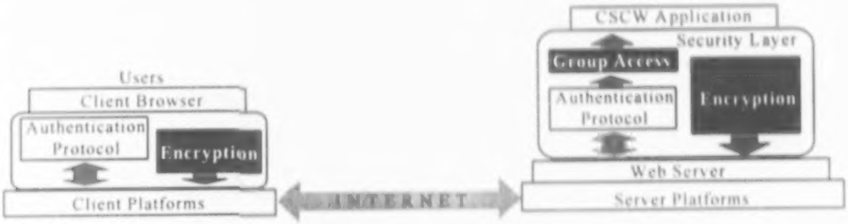


Figure 1. Secure CSCW System Architecture

References

[Bentley et al. 1995] Bentley, R., Horstmann, T., Sikkel, K., and Trevor, J. (1995). Supporting Collaborative Information Sharing with the World Wide Web: The BSCW Shared Workspace System. Schlob Birlinghoven: German National Research Centre for Information Technology.

[Doherty and Sahibuddin, 1997] Doherty, B. S., and Sahibuddin, S. (1997). "Software Quality Through Distributed Code Inspection". *Proceedings of Software Quality Engineering 97 (SQE97)*, Udine, Italy, 159-168.

[Foley and Jacob, 1995] Foley, S. N., and Jacob, J. (1995). Specifying Security for CSCW Systems. *Eight IEEE Computer Security Foundations Workshop*, June 13-15, 1995, IEEE Computer Science Press, 136-145.

[Freirer and Karlton, 1996] Freirer, A. and Karlton, P. (1996). *The Secure Socket Layer (SSL) Protocol*. Netscape Communication, Internet draft, ftp://ftp.netscape.com/pub/review/ssl-spec.tar.Z.

[Karila 1991] Karila, A.T., (1991). *Open Systems Security - an Architectural Framework*, Telecom Findland, Business Systems R&D, Helsinki, Findland.

[Koblitz 1987] Koblitz, N. (1987). *A Course in Number Theory and Cryptography*. London: Springer-Verlag.

[Sakakibara et al. 1994] Sakakibara, K., Seki, K., Okada, K. and Matsushita, Y. (1994). The ID-based Non-interactive Group Communication Key Sharing Scheme using Smart Cards. *Proceeding of 1994 International Conference on Network Protocols*, Boston, USA, 91-98.

[Shamir 1979] Shamir, A. (1979). How to Share a Secret. *Communication of the ACM*, 22(11), 612-613.

BEST COPY AVAILABLE

[Teufal et al. 1995] Teufal, S., Eloff, J. H. P., Bauknecht, K., and Karagiannis, D. (1995). Information Security Concepts in Computer Supported Cooperative Work. Norman Revell and A. Min Tjoa (ed.), In *Proceedings of 6th International Conference on Database and Expert System Application, 1995*, Berlin: Springer Verlag, 621-631.

NJECHO - AN ELECTRONIC JOURNAL FOCUSED ON COMPUTERS ACROSS THE CURRICULUM

by Beva Eastman, Ph.D.
Mathematics Department
William Paterson University
Wayne, New Jersey 07470
email: beva@igc.apc.org

The New Jersey Educational Computing Cooperative (NJECC) started aN electronic journal, NJECHO, in September 1996 . The journal is to showcase K-12 students' work in the new types of curriculum-based projects being developed using computer capabilities. This presentation will focus mainly on these new assignments leading to new learning as shown through the journal's projects.

Curriculum Areas for the 1996-1997 Year

Nine issues were published during the 1996-1997 academic year. Each month different curriculum areas were chosen as the focus for that issue. The areas chosen for the first year were:

Computer Graphics
Science K-8
Language Arts
Social Studies
Telecommunications
Mathematics
Reading
Music and Sound
World Language

The Platform of the Journal

NJECHO is distributed on one Macintosh disk at the NJECC monthly meetings. Because of the limitations of the one disk, most projects were abridged in order that several projects could be included. The presentation software used is HyperStudio™ which allows other software to be accessed. All projects submitted in other software either are converted to the HyperStudio™ format or accessed from HyperStudio™.

Format of the Journal

Each issue often used graphics from that issue to lead into a table of contents listing the projects. Each project then had its own title page with four information areas. The four areas are:
The Project - often, due to space limitations, an abridged version of the project
The Formula - the assignment was for the project
The Classroom - the grade or curriculum area in which the project was implemented
Want More Info - the teacher for the project and the name of the school with an email address or phone number, so that teachers interested in developing a similar project in their schools have a resource person to contact.

New Learning or New Knowledge Demonstrated in the Projects

The primary reason that the journal was created was to demonstrate the new types of assignments and activities that the computer allows. First, there is the ability of taking on topics that were previously introduced only to an upper level class and by using technology are now accessible to a lower level class. An excellent example of this is a projects from the Mathematics issue in which a 9th grade algebra student with the help of a spreadsheet answers the question of a standard topic in a calculus class: "what is the maximum volume that can be created given certain conditions?"

Then, there is the new experience of learning that only the technology can provide. For example in the Computer Graphics issue, 8th grade students were assigned to use two pictures: one a photograph of themselves and the second a line art of some object and then generate a "morph" between the two. In the Mathematics issue, what use to be a simple "Connect the Dots to Make a Shape" 6th grade math project to learn about coordinate graphing suddenly expanded into animating the shape. In the issue on Music and Sound, 11th grade students create their own minuets using Finale™ . Since the minuets are midi files, those files can then be used in many different situations, including the students' own web pages. The Social Studies issue showed HyperStudio™ stacks used by a large urban school district during last year's presidential election with the testing functions capabilities in HyperStudio™ stacks.

Finally and perhaps most important are the assignments that explore new forms where text, sound, and graphics are used together: the report and/or research paper such as a presentation using Astound™ on Hinduism done by 8th graders, or a second grade class, after studying the life cycle of a frog starting with live tadpoles, developed their own HyperStudio™ stack to record that cycle.

The mix of sound, text, and graphics leads teachers to thinking about the assignments differently. A group of 6th grade teachers in an Interdisciplinary unit that focuses on the Middle Ages developed the idea of a Biographical Poem. The students investigated a number of people from the period and then used HyperStudio™ stacks to publish what they've discovered. Each student chose one of the people studied and created a stack with eleven cards - each card representing one line of the poem. The lines of the Biographical Poem were: First name, Four descriptors, Relative of.... Lover of... . Who feels..., Who needs..., Who fears..., Who gives..., Who would like to see..., Resident of..., Last name.

Interesting and most exciting work is now being done in creative writing classes. Using images as a basis for expression often means that creative writing classes now make new definitions of poetry as students explore resonance between text, visuals, and sound.

Although many assignments still ask that each student develop a project, many more of the assignments are now done collaboratively, especially those where students worked with other classrooms through email and the internet: to compare scientific observations with a class in the southern hemisphere to determine the differing ideas of "daylight" or to define what is meant by culture with classes in specific regions of the world. Group projects have been usually done where a large presentation is made up of the smaller group pieces.

A particular project in the January telecommunications issue, eNJoy!, is to create and share a statewide electronic clearinghouse for information, comment, and related resources in New Jersey history, a broad-based subject common to all fourth-grade students. The project, New Jersey Online Yarns (eNJoy!), will engage 4th grade students from across the State in a collaborative "hands-on" internet experience, gathering data, interviewing citizens and documenting NJ history as seen from their community's perspective to learn not only NJ history, but also to understand the important difference between gathering information and gaining knowledge. Planned last year is detail, eNJoy! starts officially this fall.

What often is an added factor in a project is the public audience. For example, to encourage students to read, one middle school created a database of books to read. The template that was designed by the grade level staff was installed in all of their classroom computers. Students in each class entered reviews while moving through a classroom station. Classroom results were compiled, and the grade level file is now available at the library.

Each issue of NJECHO tries to find those projects which are using the computer in ways to help students become actively involved in their own learning and become responsible for their own construction of knowledge. Granted many of these projects are not new. What is new is that the technology is becoming so prevalent that teachers are showing their own creativity along with the students' creativity to learn and produce the activities. Many references for the past 10 years discuss the learning that takes place with these new technologies. NJECHO is there to celebrate the students and their teachers who are using the technology. The computer allows for many different types of experience for the student, but NJECHO tries to choose those projects that support new ways of considering and renewing previous definitions of curriculum-based activities.

The Role of the Editorial Committee

The Editorial Committee is responsible for two areas:

1) to insure the diversity of projects in each issue; not only grade level and computer skill level, but also type of school district. New Jersey, similar to any other state, has urban, suburban and rural school districts, and each issue tries to presents a range of computer software applications, a range of classrooms, and a range of school districts.

2) to produce the journal each month.

During the past year, NJECHO started to develop a web page to make the distribution more widely available. Selections from each monthly issue can be previewed, and the entire monthly issues can be downloaded. The www address for NJECHO is:

http://www.enjoy.com/njecho/njecho.html

The Future of the Journal

This next year will see some changes in NJECHO. First, the journal is to be produced on both platforms. More importantly, the journal will become similar to a monthly magazine where projects will be chosen as submitted, and the "featured" projects will possible be around a theme. either curriculum or pedagogy. This past year, most of the projects came from northern New Jersey. However, since NJECC is moving to central New Jersey, it is hoped that this move will bring projects from a wider geographical range of school districts.

Integrating Internet Technology into Distance Teaching at the Open University of Israel

Benny Friedman, The Open University of Israel, Israel, benny@oumail.openu.ac.il

Michal Beller, The Open University of Israel, Israel, michalb@oumail.openu.ac.il

Introduction

To create a more effective learning environment and extend access to a more diverse set of learners while controlling spiraling costs, the Open University of Israel (OUI) has introduced improved educational strategies and new ways of organizing its distance teaching methods [Beller 1997; Miller 1995]. The challenge is to deliver personalized, easily updated, performance-focused, learner-controlled multimedia learning tools and information to the desktop, office or the student's home, as well as to study centers. This paper will focus on the Learning Community model, and on the attempt of the OUI to integrate it with other modes of learning.

The Learning Community mode of teaching and learning functions at both the individual and the group level. It is relatively self-paced within group norms, resource-based, and occurs at different times and in different places. Interaction (student-to-student and faculty-to-student) is spontaneous [Daniel 1995]. The Learning Community mode is based on Computer Mediated Communication (CMC) which includes asynchronous telecommunications media; the creation of a virtual campus, and access to large databases, hypermedia stacks, video and text material. CMC lends itself readily to collaborative learning. A practical definition of collaborative learning is any learning activity that is carried out using peer interaction, evaluation and/or cooperation, with at least some structuring and monitoring by the instructor [Harasim et al. 1995]. In designing an on-line course, the creative challenge to the instructor is to rethink the syllabus in order to build in collaborative activities.

Computer Mediated Studies at the Open University of Israel - "Telem"

The Computer Mediated Studies project at the OUI, Telem, is an experimental project in which some 500 students in eight OUI courses are utilizing a networked environment. Electronic mail is used for asynchronous communication between students and tutors, and the submission and marking of students' assignments. Computer conferencing provides group communication, literary discourse, and interactive and reflective communication. The Internet provides ready access to libraries and resources, navigation assistance in resource searches, a bank of course-related items, and, potentially, a global network of tutors. A web site is constructed for each participating course. The site is built by the course instructor, and contains additional instruction materials, accumulated FAQs, references to relevant Internet sites and administrative data and news.

Ongoing evaluation of the project is conducted through questionnaires delivered via the WWW at the end of each semester. Results gathered from some 120 students in the 1996 Spring semester show that 70% of the students think that the CMC learning environment is more interesting than the standard one; 63%, that the environment contributes to their overall satisfaction with the course and 96% would want to take more OUI CMC courses. Two educational applications currently utilized in OUI courses are described below.

Application I: Course Updating and Collaborative Work

Today's information explosion poses two problems: dealing autonomously and collaboratively with knowledge updates requires training, and course materials must be kept up-to-date. Updating textbooks is costly, labor intensive and time-consuming. We set out to cope with both these issues by utilizing the Telem project's electronic delivery and communication nature. During the 1996 Fall semester, we implemented our approach on the Computer Science course "Computer Networks". ATM networks were the topic in focus; it was too new to have been included in the 5-year-old textbook, yet important enough already to deserve attention.

An ATM expert was invited to give an "F2F" lecture about ATM. He then joined the course's electronic discussion forum for 3 weeks to answer questions and refer students to additional resources. We suggested assignments dealing with ATM that students could carry out. Students searched for and gathered materials on the Internet; the discussion forum served as a venue for the exchange of information and negotiation of assignment topics, and e-mail was used to coordinate work. Selected projects were later added to the course Web site along with the discussion sessions, thus enriching the site with updated materials that would remain available to students in future semesters.

Application II: Synchronous IP-based Tutoring and Assistance

Though the Telem project mainly utilizes asynchronous modes of communication, we felt that synchronous modes could contribute to student-instructor interaction. During the 1997 Spring semester, an experiment using synchronous IP-based communication technology was performed on a group of high school students living in a provincial area of northern Israel and taking (in addition to their regular high-school studies) the OUI course "An Introduction to Computer Science". A 3-way Internet connection was set up (OUI, the high school and students' homes). Communication was based on video conferencing software such as VocalTec's Internet audio and conferencing products, and Microsoft's NetMeeting.

Once a week, an electronic session was held with a tutor in his office and students some100 miles away. The students were asked to study the course textbook prior to the session. Sessions began with the tutor delivering brief summaries of the material. Class assignments were then done by students, presented a few minutes later on the shared electronic white-board, followed by class discussion over the audio, chat and white-board channels.

As was expected, we found that the Internet is not stable enough for such real-time applications. Occasional Internet overloads forced us to rank the various communication channels used. It turned out that the video component was the first we could exclude in such cases. The telephony system served as a successful backup system. The chat and white-board channels were crucial, and were the least vulnerable to Internet overloads.

We concluded that standard video conferencing tools are not suitable for such educational situations, mainly because the instructor is not able to regulate the access to the communications channels, nor to override transmitted materials. Even worse, most products do not support many-to-many audio and video channels. We also found that integrating the various synchronous and asynchronous materials in an effective and natural way for students was problematic. We are actively seeking a computer application that will solve these problems for future use. In spite of these difficulties, the students did well on the final exam, and stated that they would consider taking another course only if it is delivered through an electronic learning environment.

References

[Beller 1997] Beller, M. (1997). Integrating Technology into Distance Teaching at the Open University of Israel. *http://www.aln.org/alnweb/magazine/issue1/beller.htm.*

[Daniel 1995] Daniel, S. J. (1995). The mega-universities and the knowledge media: Implications of the new technologies for large distance universities. MA thesis, Concordia University, Montreal, Canada.

[Harasim 1995] Harasim, L., Hiltz, S.R., Teles, L. and Turoff, M. (1995). *Learning networks: A field guide to teaching and learning online*. London: MIT Press.

[Miller 1995] Miller, G. E. (1995). Long term trends in distance education. Paper presented at a conference organized by the International University Consortium, University of Maryland, College-Park, MD.

Facilitating a Multimedia Design Course through Web-based Communication

Jane M. Fritz
Faculty of Computer Science
University of New Brunswick
Fredericton, New Brunswick, Canada
email: jane@unb.ca

Introduction

This paper discusses the challenges of orchestrating a course in Multimedia Design for upper year Computer Science students at the university level. The issue under consideration is to find the most suitable mechanism for: the submission of multimedia components; the submission of design journal entries; viewing submitted multimedia components by the entire class; and, providing feedback to individual students from the instructor. A web-based application to facilitate all of these activities is described.

The Course Objectives and Structure

The course under discussion was first offered in January 1997 to upper year students in a Bachelor of Computer Science degree program. The objective of the course is for students to develop an understanding of, and appreciation for, the techniques required for designing and developing an effective multimedia presentation, both for stand-alone multimedia applications and for Web applications [Fritz 1996]. This includes understanding the psychology, science, and technology of the individual multimedia components, the cognitive model of the users, design guidelines, and evaluation methods. It is hoped that the student will have developed a critical eye for effective design by the end of the course.

The initial focus on the individual components prior to any integration of components is modeled on an approach described by [Heller 1996]. We have found this to be an extremely useful approach; students are forced to concentrate on how to use one, and only one, component effectively before working at combining media. For examples, students submit assignments describing a topic using text only, using sound only, using image only. By observing their peers' submissions dealing with a wide variety of topics, the students come to appreciate that the degree of effectiveness of different media varies depending on **many** factors, including the topic at hand as well as the design concept and execution.

In this course, students have submitted a separate portfolio component to describe a chosen topic using each of: text and color; sound; image, and animation. An integrated multimedia presentation and a Web application are the two final projects. Additionally, students keep a design journal as an on-going project, evaluating multimedia presentations they seek out on their own and also evaluating submissions of their peers.

Logistical Issues

The logistics of orchestrating such a course are non-trivial:

1. Students need to have more storage space available for their assignments than for most courses;
2. Instructors need to have access to students' files for submission, and need to be able to differentiate between the submittal files and any other files that are in their course folder;
3. Instructors need to be able to view students' assignments with the same software with which it was developed, or need to be able to massage it accordingly;

4. Instructors need to be able to retrieve students' assignments in the classroom for showing;
5. Students need to be able to view the work of their peers without necessarily knowing or caring whose assignment is whose;
6. The students' design journals should be linked, when possible, to the presentations they are evaluating.
7. Providing feedback to students on all phases of their work is essential. Most of their work is viewed on-line; attaching comments directly to the work under review is ideal.

In its initial embodiment, the above requirements for this course took an inordinate amount of the instructor's time. It became obvious that the technology should be used as much as possible to help coordinate and integrate the many components of this course. The result of this realization was the design of a web application to facilitate the coordination of all aspects of this course. Weekly objectives of the course, due dates, and supplementary material were all available on the Web, so providing submission templates from the Web site was a natural extension.

The design journal had been handed in on a bi-weekly basis for feedback and marking. Students submitted evaluation sheets for each multimedia title evaluated, plus a one page critical analysis of the design features identified during the evaluation process. This input was simply changed from being paper-based to being submitted as Web forms. The resulting product of these forms is sent to a file from which can be generated further HTML files. These HTML files are generated for the instructor to view, one per student submission. They include links to each multimedia title being evaluated for which a URL has been originally submitted, and also a comment and grade field for instructor response. The subsequent submission of the instructor's comments and grades are sent to a file which students can retrieve over the Web after supplying their PIN.

A course directory has been set up with a separate folder for each student. The instructor has read/write to each student's folder. Submission of assignments is no more than the students having their files in place by the appointed time. In order to facilitate class viewing and remote viewing, the instructor originally had to spend a good deal of time in copying files to a commonly accessible area. The Web-based facility assists the instructor in creating a consistent Web interface with which to access student submissions. The resulting interface allows the instructor (in the office and in the classroom) and class (in the lab on their own time) to view the collection of submitted multimedia components in a clear, organized fashion.

A Web-based student evaluation form is available to the instructor that can be used to input comments and grade for each multimedia component assignment. This works similarly to the design journal response text fields. The instructor's comments and grade are sent to a file from where they can be retrieved by students through the Web.

Observations and Conclusions

This comprehensive Web-based course organizer provides a consistent interface for all aspects of a course that has proven to require a substantial amount of orchestrating of file accessibility. Additionally, instructor feedback is tightly bound to the contents of specific files. Having the instructor response be linked to the submitted files and design journals helps the process considerably. And, of course, the need for paper all but disappears. The techniques employed would work very well in a telelearning environment.

References

[Fritz, 1996] Jane M. Fritz (1996). Does Multimedia Have a Place in a CS Curriculum?, *APICS Conference Proceedings, 1996*, Moncton, New Brunswick, Canada.

[Heller, 1996] Rachelle S. Heller (1996). Designing a Senior/Graduate Course in Multimedia, *Workshop of Ed-Media 96*, World Conference on Educational Multimedia and Hypermedia, Boston, USA.

Virtuality of MOOs as a Language Learning Tool

Francesca O. Galante
Department of French and Italian
University of Wisconsin-Madison
fgalante@students.wisc.edu

The adoption of MOOs in language instruction spouses an approach that sees exposure to authentic language and interaction as the key-terms for language learning. As extensive studies in Second Language Acquisition have repeatedly argued, exposure to authentic language is considered the primary tool to the development of linguistic proficiency. A large part of SLA research has been devoted to investigating similarities and differences between a natural approach and an instructed approach to the language acquisition process. The results of these investigations have established very important principles in second and foreign language teaching theory although they have not been successful in transforming those principles into practice. For instance, the idea that FL/SL learners should first learn the structures of the target language and then apply them in language output was invalidated by Hatch's study in FLA in the late 1970's and by research from Sato (1988), Dittmar (1981), Klein (1981), Meisel (1987) in the following decade.

Although there is not clear evidence which supports Hatch's hypotheses for SLA, it seems reasonable that, likewise children, adult FL/SL learners recycle the linguistic assistance provided by speakers of the target language and incorporate it in their language output. The hypothesis that a non-native interlocutor may draw structural, lexical and rhetorical data from his/her conversation with native interlocutors has enormous implications for SL/FL teaching. Moreover, the study concentration on the relationship between direct exposure to language and degree of accuracy and comprehensibility in linguistic performance (Long, 1981; Parker & Chaudron, 1987; Krashen, 1980, 1981, 1982) has resulted in a shifting focus from instructed learning to a "naturalistic" approach.

We are far, here, from rejecting language instruction as we are aware that, if it has been proved the necessity of comprehensible input as theorized by Krashen, it has not been proved that a comprehensible input is sufficient to language acquisition, whereas there is a vast literature which establishes that the comprehensible input is not enough to language acquisition (Plann, 1977; Schmidt, 1981; Higgs e Clifford, 1982; Swain, 1985). It is within this perspective of a renovated trust in language instruction, provided the fact that it is accompanied by a natural setting for learners, that the notion of interaction acquires its relevance.

Language instruction around the nation has been using educational technology for at least 15 years now (Hawisher, LeBlanc, Moran, & Self, 1996) and the extended literature in favor of the use of computers to complement language learning has leaded colleges to use high-tech software. However, as Keenan (1996) points out, the costs of software are too high in terms of money and time invested in packages which soon become obsolete and are hardly retrievable in other institutions. Moreover, software packages do not promote interactive learning based on human-human communication as they only allow for a human-machine interaction. Interaction is a key-term if the language learning process rests on the idea that learners' linguistic proficiency is enhanced when they are forced to negotiate their meanings. In other words, students learn better if they feel responsible for their own comprehension and production: this will naturally lead them to manipulate the linguistic input to help their own comprehension. However, given the artificiality of the classroom setting and the limited opportunities to reach out "authentic" language speakers, the use of the Internet may promote the students' contact with a community of native speakers available to them and to the instructor at no cost. This opportunity is largely provided by MOO sites.

This paper briefly describes how an Italian MOO was incorporated in the usual practices of Italian classes, at the University of Wisconsin-Madison during the summer of 1997. During the second term of summer session 1997, students of Italian 203 logged onto the virtual environment of Little Italy (telnet kame.usr.dsi.unimi.it 4444), the largest MOO adopting Italian language created by a staff of Computer Science students at the DSI in Milan (Dipartimento di Scienze delle Comunicazioni dell'Università di Milano) at the University in Milan.

As a second year of language instruction, Italian 203 sees the development of oral and writing skills already as a priority toward the linguistic proficiency that students will need in the upper courses. Exposure to the authentic language is considered the primary tool to develop that proficiency. By bringing Little Italy in the Italian classes students would gain much more than an intriguing new source of information: they would get access to Italian speakers (mostly native) around the world and in Italy, and they would interact with them. Internet offers a large number of MOO sites in different languages, however very few are in Italian. Italian language instruction in most North American academic institutions is largely tied to audio lingual practices as most of the textbooks and syllabus (mostly grammar-oriented) witness. The integration of a text-based virtual community in the language class presented a number of pedagogical advantages.

Before introducing the MOO in the language class, the students were made aware of the notion that behind each of the keyboards, wired into the networks, there was another human being who might be logged on to the MOO database. Any of those remote users might become their language (writing) partner in real time. As it turned out to be, most of Little Italy users shared age and occupation with the American learners of Italian: communication with the overseas speakers, therefore, was enhanced by both similarities in interests or in academic goals and differences in culture and lifestyle. The American students experienced a great deal of curiosity and welcoming attitudes by their Italian counterparts, also due the noticible massive introduction of 19 foreign players in the MOO. It must be added that the DSI staff offered an invaluable help for having promptly assigned each student with a character and even offered itself as interlocutor to the class during the scheduled class visits to Little Italy.

Thanks to the educational technology resources available at the Learning Support Service department at UW-Madison, the students were involved in collective MOOing sessions throughout the duration of the course, during which they were gradually introduced to the basics of the Lambda programming language and overcome their eventual computer illiteracy. In addition, they were accompanied to tours of the virtual city, and performed as guests their first text-based conversation with their distant partners.

Since the students' access to Little Italy was usually provided via a telnet client, at first they encountered some difficulties due to the telnet features. Those same features, however, constrained a linguistic behavior which had a great impact on the learners' attention to their accuracy and brought them to control their linguistic output: they found that spelling and grammar imperfections might have interfere in the transaction of meanings with their interlocutors and, for the sake of their role in the interaction, they would pay extraordinary attention to what had been said. For the first time grammar and spelling was not an annoying part of the language usually responsible for their course grade, but a real tool of communication. At the same time, however, the students experienced inefficacy of linguistic accuracy in communication if this was not accompanied by linguistic appropriateness. Furthermore, the text-based communication strengthened the students' awareness of the absence of nonverbal communication cues and led them to understand on their own how important precise word choice becomes in an environment devoid of audible tone and visible facial expression. The text-based communication, in other words, established a need for grammar, spelling and lexicon accuracy, as well as a need of expressing meanings in appropriate manners and the importance of conversational strategies (Haynes, 1995). The MOO structure is strongly oriented to socialization. This feature enforced the learners of belonging to a linguistic community and, therefore, it committed students' engagement in conversations with highly communicative tasks. In few sessions the students became learned about strategies to start a conversation with another MOO user, to keep a conversation alive and to maintain it attractive. Such strategies involved a continuous occurrence of paraphrasing of students own utterances, which in turns helped to increase the learners' awareness of their pragma-linguistic competence.

The language teacher may perceive the pedagogical advantages in the students' behavior above described: the MOO allows for the reflection on one's own linguistic output without affecting the spontaneity and veridicality of the linguistic and communicative competence. By having a control on their own language output and by reflecting on the language produced during the interaction, learners take responsibility for their own learning. Learners are offered the opportunity to explore their linguistic competence: to play with structures, to learn them from their native speakers interlocutors, to try them in the interaction. All this without the affective constraints of the classroom setting.

Finally, the fictional status of the futuristic city of Little Italy well served the goals of the conversation an composition courses in Italian: the students were induced to create their own persona, environment, and belongings by using a fictional descriptive language; in the interaction with the other players they used expressive and argumentative language; in reporting to the class their own life as a fictional characters, they translated their MOOers' experience into narrative.

The integration of a virtual reality system in the traditional approach to language instruction, then, seems to present valuable pedagogical advantages: as a low-cost technological facilitator of the students interaction with native speakers, the MOO also allows the learners to develop their own evaluation of a communicative situation and to connect linguistic competence to pragmatic competence as they are involved in a decision making process about how to act, to react, and to interact.

References

Dittmar, N. (1980). On the Verbal Organization of L2 Tense Marking in an Elicited Translation task by Spanish Immigrants in Germany. Studies in Second Language Acquisition, 3 (2): 136-64.

Hatch, E. (1978). Discourse Analysis and Second Language Acquisition. In Hatch, E. (ed.) Second Language Acquisition: a Book of Readings. Rowley, Mass.: Newbury House. 402-35.

Hawisher, G. H, et al. (1996). Computers and the Teaching of Writing in American Higher Education, 1979-1994 : a History. Norwood, N.J. : Ablex Pub.

Haynes, C. (1995), Synchroni/City: Online Collaboration, Research and Teaching in MOOspace. http://lingua.utdallas.edu:7000/748

Higgs, T. & Clifford, R. (1982). The Push toward communications. In Higgs, T. (ed.) Curriculum, competence, and the Foreign Language Teacher. Skokie, Ill.: National Textbook. 57-79.

Keenan, C. (1996). Technology in English 015: Building Low-Cost, High-Powered Writing Communities. http://cac.psu.edu/!cgk4/horizon.html

Klein, W. (1981). Some rules of Regular Ellipsis in German. In Klein, W. & Levelt, W. (eds.) Crossing the Boundaries in Linguistics: Studies Presented to Manfred Bierwisch. Dordrecht: Reidel. 51-78.

Krashen, S. (1980). The Input Hypothesis. In Alatis, J. (ed.) Current Issues in Bilingual Education. Washington, D.C.: Georgetown University Press. 168-80.

Krashen, S. (1981). Second Language Acquisition and Second Language Learning. Pergamon.

Krashen, S. (1982). Principles and Practices in Second Language Learning. Pergamon.

Long, M. (1981). Input, Interaction and Second Language Acquisition. In Winitz, H. (ed.) Native Language and Foreign Language Acquisition. Annals of the New York Academy of Sciences, 379: 259-78.

Meisel, J. (1987). Reference to Past Events and Actions in the Development of Natural Second Language Acquisition. In Pfaff, C. (ed.) First and Second Language Processes. Cambridge, Mass.: Newbury House. 206-24.

Plann, S. (1977). Acquiring a Second Language in an Immersion Situation. In Brown, H. et. al. (eds.). On TESOL '77. Washington, D.C.: TESOL. 213-23.

Parker, K. & Chaudron, C. (1987). The Effects of Linguistic Simplifications and Elaborative Modifications on L2 Comprehension. UHWPESL, 6 (2): 107-33.

Sato, C. (1988). Origins of Complex Syntax in Interlanguage Development. Studies in Second Language Acquisition, 10 (3): 371-95.

Schmidt, R. (1981). Interaction, Acculturation and the Acquisition of Communicative Competence. University of Hawaii Working Papers in Linguistics, 13 (3): 29-77.

Swain, M. (1985). Communicative Competence: Some Roles of Comprehensible Input and Comprehensible Output in its Development. In Gass, S. & Madden, C. (eds.) Input in Second Language Acquisition. Rowley, Mass.: Newbury House. 235-53.

MULTIMEDIA PEDAGOGY - Creating longevity in CAL applications !

"[Multimedia applications are] like sex. When all is said and done more is said than done."

J.G. Gallagher,
Napier Business School
Napier University
Sighthill Court
Edinburgh
Scotland
j.gallagher@napier.ac.uk

E. Fordyce
Thames Valley University
St Mary's Road
Ealing
London W5 5RF

D.P. Stevenson
Napier Business School
Napier University
Sighthill Court
Edinburgh
Scotland
d.stevenson@napier.ac.uk

Although the above mis-quotes Joseph Heller. It allows, nevertheless, an insight into the current status in the development and production of multimedia applications. Simply put, sex sells! Sex grabs the imagination! It offers the allure of hidden promises and unfulfilled desires! However, when sex or, more properly, sexiness is applied as the vehicle for promoting and selling multimedia applications it inevitably falls short in delivering its promises. It fails to achieve the consummation of the marriage between platform and content. It fails, once used, to fire the soul. It fails to maintain a lasting relationship with the user. What the user ends up with, more often than not, is multimedia applications dressed to kill, all 'glitz' and no content. If multimedia applications are to deliver their promise then content rather than platform must be the primary driving force. It must fully address the end user's needs and wants. It must provide substance.

The key, therefore, to the development and application of multimedia is an emphasis on quality, but quality not just in the delivery platform and structure but in content and outcomes. All too often the subject content is undemanding, inexpertly explained and apparently without a clear concept of learning outcomes. In these cases multimedia may be seen simply as a means of adding 'glitz' to a teaching programme through the creation of a more colourful and animated delivery system. If, on the other hand, the aim is to increase user knowledge and understanding then the quality of the content is crucial. Where there is an emphasis on content quality, manipulation of multimedia elements can achieve the integration of different content components more effectively than is possible on the printed page. This marriage of content elements should be a major target in the process development of multimedia applications.

New applications could, therefore, be examined in the cold light of acceptable pedagogical and user interface dimensions (Reeves and Harmon1995) e.g.

Pedagogical	*User Interface*
epistemology	ease of use
underlying psychology	navigation
experiential value	screen design
value of errors etc.	media integration

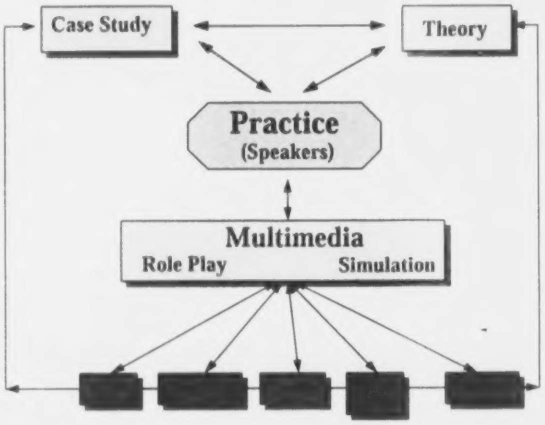
If these dimensions were applied as criteria against which to measure user satisfaction or effectiveness then they would allow some measure of success or failure for multimedia applications to emerge. The development of thorough conceptual understanding involves a series of learning phases (G.J. MacFarlane 1995) - preparing to tackle the relevant material, acquiring the necessary information, relating it to previous knowledge, transforming it through establishing organisational frameworks within which to interpret it, and so developing personal understanding (CSUP Report 1992). If this process is to work effectively, teaching -however it is delivered- must be designed to support these phases of learning. The required support can be described in terms of necessary teaching functions which to some extent parallel, but also overlap, the phases of learning. These functions include:

orientating -setting the scene and explaining what is required
motivating - pointing up relevance, evoking and sustaining interest
presenting - introducing new knowledge within a clear, supportive structure
clarifying - explaining with examples and providing remedial support
elaborating - introducing additional material to develop more detailed knowledge
consolidating - providing opportunities to develop and test personal understanding
confirming - ensuring the adequacy of the knowledge and understanding reached.

The implication for multimedia presentations maybe that the applications design is reversed from what is currently the norm. The apparent trend of building an application because we have a platform would be subsumed by the necessity of providing quality content geared to satisfying the end users' needs.

Traditionally, on MBA courses the case study method would be used as the vehicle to develop student learning and understanding as the case study facilitates the marriage of theory and practice. This would then often be augmented by having a guest speaker from the case company to address the class. The objective is to structure the case so that it accommodates crucial areas of :

level (undergraduate; postgraduate; post-experience etc.)
complexity (stage of the course it is intended for - introduction or final examination case.
currency (what is the shelf life of a case - one year or five years?)
target group (accountants; general business students or behaviourists etc.)
outcomes (what is to be achieved by using the case - academic; social; etc.?)



BEST COPY AVAILABLE

Figure 1. highlights this relationship. Multimedia techniques appears to offer a tremendous potential for, not only achieving the objectives of case development but, augmenting them. Consequently, examination and evaluation of this potential was undertaken. The application of multimedia techniques with their inherent flexibility appeared to offer the best potential for ameliorating some of the problems associated with the use of case studies e.g. not all students learn at the same rate, nor do they start from the same educational base - in the area of business policy in particular they are likely to come from a range of disciplines, nor are all students as ready to contribute to class discussion.
Where *figure 1* highlights the traditional relationship between case, theory and practice the addition of multimedia elements allows the introduction of additional material in a more student controlled environment (See *Figure 2*).

The system should aim to test all students whilst allowing each student to progress at his/her own pace. It should promote incremental development of problem solving skills and increase the effectiveness of learning. Students should learn (partly by doing) to organise their own work patterns and determine the means to overcome the difficulties associated with solving complex, unstructured problems. The mere adaptation from paper to text on screen will not achieve these.

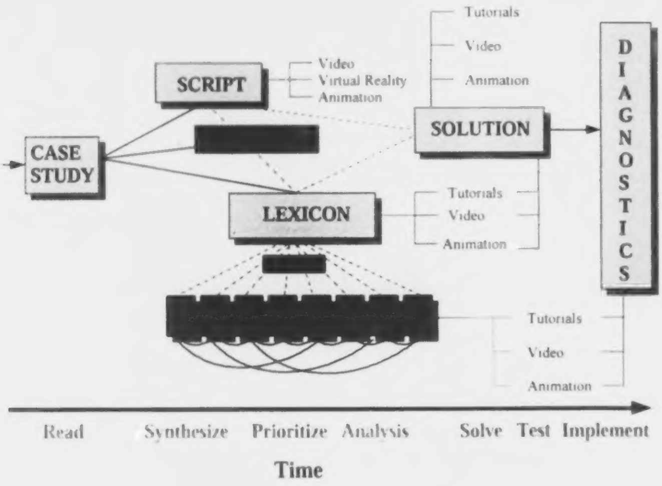


Figure 2. Flow Chart

However, it would be fair to say that the multimedia case is using computers to do what cannot be done either on the printed page or with blackboard and chalk. There can be little substitute for the symbiotic development of ideas and solutions generated by students in a lecturer led, class based, case discussion. There is no definitive solution to any given case study. There are though, a number of routes to a number of possible solutions. The interactive case has the advantage that it can present to the student what the company actually did and the rationale that lay behind its decisions whilst still allowing the student to explore other options.

From the flow chart it can be seen that the key elements to be built are

- *case study* - video, scripting
- *script* - video, sound, graphic design, animation, virtual reality interface, hypertext links
- *lexicon* - video, dictionaries of theory, tutorials, self assessment
- *solution* - video, applied theories, system interrogation
- *simulation* - applied learning by competing against other industry competitors

all of which must be supported by the software development of:

- *software applications* - hypertext scripting
- *fractal compression*
  - *virtual reality construction*
  - *scripting library*
  - *lexicon library*

BEST COPY AVAILABLE

Subsumed within this knowledge of what to build is a deeper knowledge of what producers of interactive computer aided learning should attempt to satisfy. Zimmerman (1989) viewed the CD ROM as a vehicle to provide learning through guided experience. Figure 3 crudely attempts to link outcomes with the potential requirements demanded of the product. Essentially, it tries to show that the key to good CAL lies in providing strong guided experience.

The System: should be robust, easily navigated, simple to control and fully interactive. It should facilitate the transfer of knowledge without requiring the user to develop computing software skills.

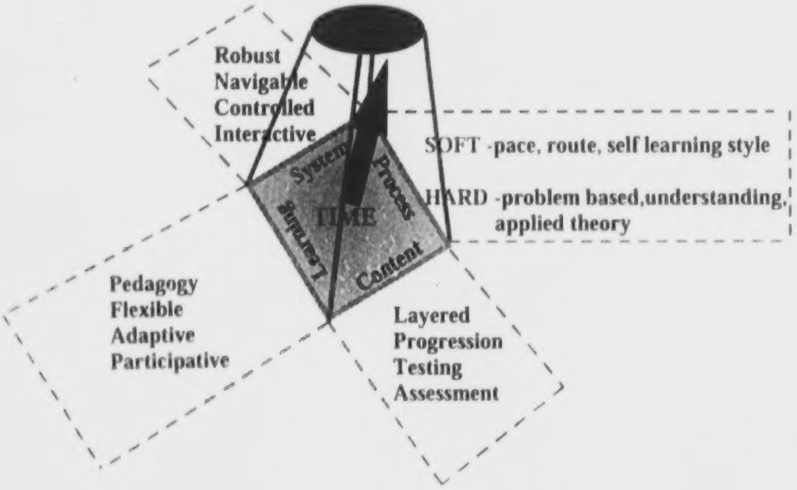
The Process: The process elements may be viewed as either soft or hard. The soft elements are those which the user should have some control over e.g. the pace of learning, the route which best suits his/her needs and the ability to accommodate the individuals own self learning style. The hard process elements acknowledge the requirements to produce problem based learning processes, user understanding of the process dynamics, and the process interface when developing applied theory. In reality it is probably the case that the hard elements far outweigh the soft ones which may in reality be more perceived than real.

Content: should be layered one level on top of another. User progression from one layer to the next, to some extent will be dictated by a combination of both the soft and hard process elements for example, a user may decide that his/her knowledge base is sufficient enough in a given area that work in this area can safely be avoided. In this instance he/she has made the decision not the system. However, at a later stage the system testing will assess whether he/she has the knowledge and understanding to adequately ignore this section and on the basis of this recommend appropriate action. This can either be done by the system setting tutorial assessment on that area alone, before progression is sought, or later when random tutorial selections are made. In any event, the system will assess content understanding.

Learning: is predicated on the system having the appropriate pedagogical input which is both flexible and adaptable. It should underpin the system content and should be adapted to user participation.

Time: is often forgotten in the development of CAL applications. Time is needed by the user to learn the system, its navigation, its processes, and its interface. Understanding the impact of time is crucial in construction phase of these applications. For example, it is unlikely to be appropriate to simply take the sequence of learning, as indicated by a curriculum and superimpose the curriculum content on to a CD ROM.

The essence of CAL, to a great extent, is to pass over the sequence of learning to the user when learning is freed from the confines of the classroom. But, without guided experience success in terms of learning is far from assured.



BEST COPY AVAILABLE

Preliminary testing of the CD ROM was undertaken on the MBA Programme at Napier University and will continue on a wider scale in the coming session. Initial results from questionnaires, and from individual interview are limited therefore in their applicability but they did throw up a number of observations which should be examined.

To begin with, the students found that the CD ROM allowed them to adapt their learning style. Essentially, this meant that each student created his/her own personal learning style by redefining their routes to learning and knowledge acquisition. The nature of the original case medium meant an holistic environment was presented to the student which represented a complex unstructured problem. The resolution(s) of this problem allowed the student to develop and test skills and techniques in a more challenging environment, one moreover, which provided learning by doing through its iterative process. Control of this process was seen by the students as being a positive feature of the learning process. One moreover, which allowed them to exploit time. They were no longer tied to taking notes in class but were free to roam through the linkages they wished to explore. They could ask questions and seek the answers at the pace and time dictated by themselves.

The lexicons were used by the students as dictionary bases which both presented and clarified new knowledge. However, it was apparent that the lexicons were also being used to support studies in other areas than simply corporate strategy. They were being used as supplements for the finance and marketing courses. One factor which emerged associated with this was that the students were making suggestions on the need to augment the hypertext links between individual lexicons to allow a freer access to additional data bases.

The provision of the worked solution was also highlighted by the students as beneficial. This was viewed as a means by which they could introduce theory garnered from other sources to be applied to the problem posed. Moreover, the introduction of video of the personnel who actually faced the situation posed in the case and the rationale given by them for the decisions they took brought elaboration of a more forceful nature.

The tutorial application was highlighted by the students as a consolidating function. But, they displayed a resistance to yes/no; true/false type questions. They preferred multiple choice questions. However, the system's facility to randomly select questions from its bank and to provide worked solutions on request which allowed the students to test themselves in a more forgiving environment than the classroom was seen as a positive application especially when liked by the systems revision feedback.

A further development which students saw as supporting the learning process was the diagnostic tools embedded in the system. These provided a platform to aid student decision making from a practical basis as they support scenario planning utilising the student's own information input. This became important to the student when it became obvious that the traditional compartmentalisation of subjects could be broken down when the diagnostic tools were used to support other subject areas.

If we are to satisfy the end user then today's quantitative output of multimedia applications must become tomorrow's qualitative output where quality content is best exploited by a quality platform. There needs therefore, to be a re-examination of both the purposes and the techniques involved in the construction of computer aided learning (CAL) applications. Longevity in Cal applications can only be maintained by providing products which are robust technically as well as challenging intellectually.

References

Reeves and Harmon1995	Sorel Reisman, Multimedia Computing, Preparing for the 21 Century Idea Group Publishing Harrisburg, USA
G.J. MacFarlane 1995	Future Patterns Of Teaching And Learning Heriot-Watt University Edinburgh Scotland, UK.

CSUP Report 1992	CSUP Report (1992) Teaching and Learning in an Expanding Higher Education System, Committee of Scottish University Principals Working Party (Obtainable from COSHEP, St Andrew House, 141 West Nile Street, Glasgow, Scotland, UK.)
Zimmerman (1989)	Brown Collins and Duguid,1989

MULTIMEDIA PEDAGOGY - Creating longevity in CAL applications !

Abstract: Pedagogical content is fundamental to developing longevity in multimedia educational applications. All to often producers of such applications are captivated by the technology of the delivery system, producing sparkling, colourful presentations with little regard to their content. This, however, inevitably leads to a one-sided, ineffective application. Multimedia is the singer not the song. It stands or falls on the quality of its content. It is a tool which when combined with the right content provides a teaching and learning vehicle which significantly contributes to the learning process. This paper traces the experiences of developing electronic, interactive, multimedia, for use on part-time MBA and Distance Learning MBA courses. It attempts to evaluate the sequence of learning traditionally undertaken by students and juxtaposes this with the sound-bite, self service learning mode offered by multimedia.

The paper attempts to explore some of the pros and cons encountered by the student in this learning process and to draw from these instruction on construction of computer-based learning (CBL) elements such as the interface between delivery platform and content and how the integration of these relate to the need to develop new client/server implementation and infrastructures. In short, it questions the efficacy of current claims for multimedia of providing CBL to augment orthodox teaching and training. In particular, it questions the claim that it is a means of providing a realistic, seamless, integration of text, sound, pictures, graphics, animation and hypertext. It holds that this new learning process can only be predicated on the assumption that new applications which stand the test of time and usage are based on content quality Necessarily, the paper will be supported by a demonstration of the system being developed by the authors.

J. G. Gallagher
D. P. Stevenson
E. Fordyce
Napier Business School, Napier University, Edinburgh

Brokering the Web

K. Garcha
I.N Research Team
GPT Limited
U.K
kul@ncp.gpt.co.uk

Introduction

This summary outlines, what Information Brokering is and discusses three important issues involved in realising this concept:

query processing,
automated negotiation,
the integration of external packages into a user's working environment.

The work which contributed to this paper was conducted as part of a U.K. collaborative project VIRTUOSI [Virtuosi, 1996]. involving industry and academic institutions.

Information Brokerage

The increase in the number of commercial information services (information, software applications, entertainment, etc.) which are publicly accessible from the world's networks, has brought with it its own problems. Information service providers need to effectively market their products and services in a totally new way. A new automated trading environment is developing which requires new management and administration mechanisms. User's are faced with access to an ever increasing mountain of information but do not have the tools capable for efficiently locating, processing and managing this information, causing the problem of information overload.

One way of tackling these problems in an integrated manner is by using an intermediary called an "Information Broker".

Negotiation

Negotiation is a joint decision making process in which various parties state their requirements, some of which may conflict, negotiation allows all parties to move towards agreement by a process of concession or the search of new alternatives. Negotiation relies heavily on the ability of the brokers and agents to communicate and to understand each other. Messages need to be standardised by building common ontologies, message wrappers etc. There can be various types of multi-agent negotiation depending on the type of environment the broker is dealing.

Automated negotiation is currently in its early stages, and there is no clearly defined interaction protocol enforced by law since different countries may be governed by different laws (enforcing such laws may be impractically expensive). Since a computer agent can vanish at any point in time, laws can only be enforced if the terminated agent represented some real world party and the connection between the two can be traced. A broker is in an ideal position to enforce an interaction protocol by acting as a trusted intermediary, making sure any laws are adhered to and any other business requirements such as accounts, taxes are also met.

Integration

One of the other areas to which a broker can greatly contribute is by speeding up the complicated task of integrating external applications into a multi-user virtual working environment (e.g. wordprocessors, games, cad tools, cost models etc.).

A multi-user virtual environment allows applications to be used interactively for remote team based activities enabling real time arbitration, communication and co-operation. Each user can edit the contents with each action being reflected to all the other users. In the case of a cost model, a team of designers of a product, located in different places could meet in their virtual work environment and interactively collaborate.

There are many issues (e.g. network technology , I/O devices, performance) that need to be considered when integrating such applications. Tools that make integration easier are emerging.

The broker could also handle administration and disintegration issues. Having decided that a particular piece of software can be integrated. The broker could handle the production of the contract i.e. licensed use of the software, upgrades, payment details etc. On expiration of the licensed date, a mechanism is needed to disable the software from the system and clear up any payments due.

Query Processing

The broker needs to efficiently process queries from users, retrieving and integrating distributed data can be very costly. Query processing involves developing an ordered set of operations for obtaining a requested set of data, such as selecting the information sources, choosing operations for processing the data, selecting sites where the operations will be performed and the order in which they will be performed. What is required is an automated dynamic system to generate and execute query access plans. This planner needs capabilities such as executing operations in parallel, re-planning queries that fail while at the same time executing other queries, gathering additional information to aid the query processing and the acceptance of new queries while other queries are being executed.

Conclusions

This paper has discussed the issues of automated negotiation, integration and query processing which are seen as being crucial to the realisation of Information Brokerage.

References

[Virtuosi,1996], Virtuosi Project, http://www.crg.cs.nott.ac.uk/Virtuosi.

[Garcha, 1996], Garcha. K. (1996). Realisation Issues for Brokerage, Virtuosi.

[Foss J, 1996], Foss. J. (1996). Brokering the Info-Underworld, UNICOM Seminar on Real-World Applications of Intelligent Agent Technology, London, 1996.

EDU-EX: Design of a Knowledge Based Tool to Create Pedagogical Intelligent Educational Systems in the WEB.

Angel García-Crespo, Paloma Domingo, Miryam Lancha, Belén Ruiz Mezcua, Antonio de Amescua
Computer Science Department, Universidad Carlos III de Madrid
Butarque 15, 28911 Leganes (Madrid)
Tel: 34-1-6249417, Fax: 34-1-6249430
E-mail: agarcia@rioja.uc3m.es

Introduction

We can define an Intelligent Teaching System as the software system that can be adapted to any student situation and has control over the system itself. The ITS also allows the development of interactive courses simplifying transmission of knowledge from human experts to others without impediments.

The Intelligent Teaching System must be able to detect and control the characteristic properties of each student ; such as, the capacity to absorb knowledge, frequency and the rate of study, response time, changing rates in response to different factors, etc. At hand is not just the detection of these properties, but also diagnoses of modifications and the reasons for this. Hence, depending on the diagnosis, we can evaluate the needs that each student has during a training period. This diagnosis mechanism has to evolve and modifying continuously for it to be able to be adapted to the learning rate of individual students.

The Intelligent Teaching System has to detect the level of understanding possessed by students in specific subjects. This level will be a function of the course itself and the properties the student itself introduces during the course. The Intelligent Teaching System must be capable of designing the most suitable didactic strategies at all times. It should know how to choose the right moment for revising a lesson considering real data as the study rate, the complexity of the contents, how often a concept was explained, etc. Considering what we know up to now, the system has to be capable of adapting to the student's characteristics by the use of dynamic mechanisms.

EDU-EX was developed in WEB environment in order to facilitate the delivery of educational materials all over the World and to virtually every current platform.

The intelligent tutoring system was created PC windows environment. The course was developed by experts (domain, educator and programmer). When the course is finished it is stored in the server, as the diagram shows. Different students can use this course by simply connecting into the web pages on the course. Its characteristics are stored in the server facilitating one-to-one teaching. In this way, we can overcome the main problems associated with more classical systems. Each student link onto his/her own personalised intelligent tutoring system all over the world and within current platforms.

The interaction of EDU-EX is shown in the diagram below:

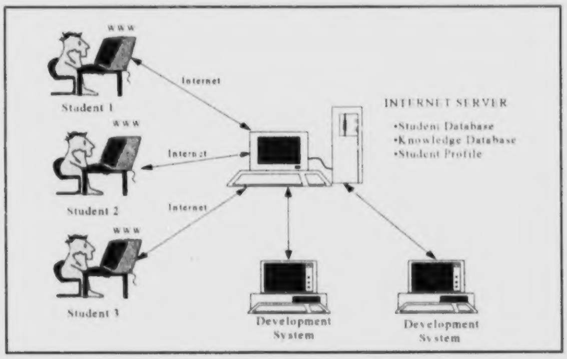


Figure 1: System Implementation

BEST COPY AVAILABLE

System Development Tool.

EDU-EX is a system that allows the creation of Intelligent Educational Systems in the WEB based on decision support systems (expert systems). Any area, the result of a situation in the real world, can be solved with EDU-EX, making decisions that are adapted to each specific area and verifying the decision taken on the area and in relation to the decisions taken previously. EDU-EX uses objects to organise the information in the knowledge database. The knowledge is stored in the properties of the objects. The most important objects of the EDU-EX structure are the AREAS. The rest of the objects such as; DECISIONS, ACTIONS or TESTS serve as support to AREAS objects.

AREA objects allows to create a network that contains domain knowledge to be teached.

Pedagocical planification is accomplished in two diferent ways:

Using an object named GUIA that allow us to change AREA network dinamically (considering student performaance and answers to tests) and/or using another object named DECISION that permit to change the way in which network searching is done. Both objects GUIA and DECISION are the objects that allow pedagogical planification with no constrain.

The Area Solver is the module that processes information stored previously in the knowledge database. Using these areas the area solver perform decision-making management tasks.

Primarily, the Area Solver is responsible for: the description of causes that give rise to the area. The suggestion of decisions area solving, which are adapted to the decisions each user requires at a specific moment. Verification that decisions have been taken correctly and that solutions have been found for the area. EDU-EX is a non monotonic system. This implies that during a specific session , the user can modify the value of a response given previously. The Area Solver automatically restarts process of reasoning, considering the results of solutions already tested and the solutions abandoned during the process.

Conclusions

The Intelligent system proposed is another component to be included in Multimedia Intelligent Systems. It can be used either as a basic support for correct teaching or as an additional element that make stronger others teaching tools.

Currently we are ending the development of the tool in Windows 95. The tool has been proven with several real systems obtaining very good results. The accomplishment time of the phase of knowledge acquisition has decreased of 200 hours, in a traditional expert systems tool to 10 hours. The knowledge representation and area solver use an equal model used by a human expert. The study of the knowledge model representation as areas tree and of the strategies of the decisions, they are the fundamental aspects of the system.

References

[Adar94] Adarraga, P. y Zaccagnini, J.L. (1994). *Psicología e Inteligencia Artificial.* Madrid: Trotta S.A.

[Mate88] Maté Hernández, J.L. y Pazos Sierra, J. (1988). *Ingeniería del conocimiento: diseño y construcción de sistemas expertos.* Córdoba, Argentina: Sociedad para estudios pedagógicos

[Wate86] Waterman, D. A., (1986). *A guide to expert systems.* Addison-Wesley Publishing Company

[Harm85] Harmon P., King D. (1985). *Expert Systems.* John Wiley & Sons, Inc.

[Haye83] Hayes-Roth F. D., Waterman A., Lenat D. (1983). *Building Expert Systems.* Addison-Wesley.

[Ade92] Anderson, J. (1992) *Intelligent Tutoring and High School Mathematics.* Intelligent Tutoring Systems, 2nd International Conference.

[Grub94] Gruber, T. (1994). *NIKE : A National Infraestructure for Knowledge Exchange.* Proceedings of the NIST/AED Workshop on Learning Technologies.Holland, J. (1993) *Echoing Emergence: Objectives, Rough Definitions, and Speculations for Echo-Class Models.* Report 93-04-023. Santa Fe, N.M. : Santa Fe Institute.

[Kapl93] Kaplan, R., H. Trenholm, D. Gitomer, and L. Steinberg. (1993). *A Generalizate Architecture for Building Inteligent Tutoring System* in Proceeding of Applications of Artificial Intelligence

[Weng87] Wenger, E. (1987). *Artificial Intelligence and Tutoring Systems.* Los Altos, California.

Directory of Available Projects:
A Tool for Students, Faculty, and Administrators
htttp://www.WPI.EDU/~trek/webnet97/

Joanna L. Gaski, College Computer Center, jgaski@wpi.edu, Worcester Polytechnic Institute, MA, USA

Amy L. Marr, George C. Gordon Library, trek@wpi.edu, Worcester Polytechnic Institute, MA, USA

Charles Kornik, Projects & Registrar's Office, cjkornik@wpi.edu, Worcester Polytechnic Institute, MA, USA

Introduction

The WPI Projects Program is a unique facet of WPI's highly reputable undergraduate program. Each student must satisfactorily complete three large projects, a sufficiency, an interdisciplinary qualifying project (IQP), and a major qualifying project (MQP), to receive an undergraduate degree. Faculty members and students are encouraged to design projects, and the Projects and Registrar's Office is responsible for reviewing and displaying proposals.

Until last year the project selection process was disorganized. Faculty members would type up proposals and post them outside their offices, or send them to the Projects Office, which would place a three-ring binder of proposals in the library. Other departments would publish projects booklets, or put up small web sites of department project information.

Webmaster was asked by the Projects Administrator to organize the project proposal process. For this purpose, we created the Directory of Available Projects (DAP), at http://www.wpi.edu/Academics/Projects/available.html, which enables the submission, review, and browsing of project proposals. A proposal includes information such as the code, title, topic area and description of a project, and the name, email address, phone number, office and department of the advisor.

Design & Implementation

The functions of the DAP and the groups which can use each are listed in the following table.

Function	*Group(s) Involved*
Submitting Proposals	Faculty
Accepting/Rejecting Proposals	Projects Administrator
Viewing Accepted Proposals	Students, Faculty
Claiming Projects (Deleting Proposals)	Faculty, Projects Administrator

Table 1: DAP Functionality

Faculty and staff input descriptions of the projects they are proposing into an HTML form. This form calls a CGI script which separates the input fields, performs some input validation (empty fields, duplicate project codes), notifies the Projects Administrator of the proposal by email, and sets the proposal in HTML (DTD 3.2). Professors' email addresses are set as mailto links for students' convenience in contacting them. The HTML created for each proposed project is sent to a directory of proposals to be reviewed.

Project administrators can view the descriptions of each of the submitted projects awaiting acceptance. A project can be accepted, rejected, rejected without mailing, or considered later. A professor is automatically notified by email when his/her project is reviewed. This mail includes a text reason, input by the administrator via the web, if the proposal is refused, and that proposal is removed from the web site. Accepted proposals are moved to the appropriate directory of our web server for projects of that type, based on the HTML topic code. Access to the script which does this is controlled using the .htaccess and .htpasswd facilities of the Apache web server.

Students can select the topics or majors that interest them from an available projects page, and another script presents only those projects that match their selection. Students can browse for projects in their fields in this way, and email professors (by a mailto link) requesting the project or asking for more information

After a group of students has claimed a project, administration can delete that project from the available area. A Perl CGI script moves these projects to another area of the DAP which is not publically accessible. The functions of the DAP are illustrated in the figure below.

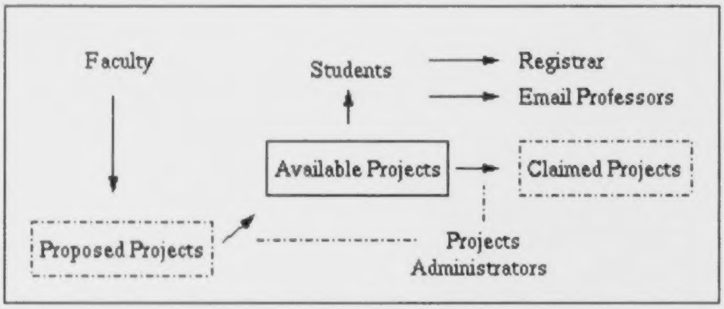


Figure 1: The DAP Process

Results & Future Work

Over 200 projects were submitted by 107 faculty members by mid-February. The DAP saw 1000 hits over the course of the month of February. No paper related to faculty project proposals was generated by or received from the Projects Office.

Faculty and students have found the DAP "easy to use." They have also found it "well organized." Faculty were also pleased that they did not need to know HTML in order to use the DAP. Unfortunately, some departments stuck to their traditional method of information distribution, which led to some confusion on the part of the students.

Based on community feedback, there are several features we would like to add to the DAP. Among these are a two-step submission process that would allow faculty members to see the HTML generated for a proposal and to correct errors before submission. Another change would enable corporate sponsors to submit project proposals. A partner location area for students who have chosen a project but who do not have a full team would be another addition. Finally, we would like to allow faculty to edit and delete previously submitted proposals.

References

[Merlic, Craig & Matthew Walker, 1996] *Virtual Office Hours: A Communication Tool for Students and Teachers.* WebNet96 Proceedings Online. http://aace.virginia.edu/aace/conf/webnet/html/317.htm. Their work in *Virtual Office Hours* at UCLA (http://www.chem.ucla.edu/uclavoh/) inspired us to create this site, and their Perl scripts formed a foundation for our projects area.

Creating an "Exchange" Through Online Web-based Events

Melanie Goldman
National School Network
BBN Corporation
70 Fawcett Street
Cambridge, MA 02138
mgoldman@bbn.com

Introduction

Can we create a space on the web that brings excitement to the classroom, encourages interactive dialogue, and stimulates collaboration? A most promising innovative and participatory web-based model has emerged in the National School Network's web-based "Exchange". What we have been able to do through the web has been phenomenal in opening up the school walls to bring in the world and to kick start use of the Internet in the classroom. Key to the "Exchange" are a wide range of mechanisms such as online events, case studies of member experiences, newsletters, tools, cookbooks, as well as the emphasis on building partnerships with local businesses, schools, and community organizations.

Background

The "Exchange" has been under development since 1992, sponsored by the National Science Foundation as part of the National School Network (NSN). This work evolved from a combination of BBN technologies in Internetworking and client/server architecture, seminal research with local area networking in schools in the late 1980's and early 1990's, and a set of premises about educational reform in a networked environment. The NSN is a community of over 500 schools, school districts, museums, universities, businesses, and government agencies which are leaders in integrating the Internet into the curriculum in support of educational reform. To achieve educational benefits from networking, the NSN and its members are inventing new kinds of educational activities. It has been the experience that in the current, early, stage of building local school use of the Internet, the most promising applications are those which directly engage students in learning with and from other people, such as special events with experts and telementoring.

Online Events

What has been particularly successful is a series of exciting online events where students have the opportunity to talk directly with contemporary figures who are exploring scientific phenomena, making policy in government, authoring stories, or journalists from nationally syndicated newspapers. For example in a videoconference with the

NASA astronaut Dr. Dan Barry, students were able to ask him, "Do things grow in space?" "What is it like to wear a space suit?" and "What types of experiments did you conduct while in space?" Dr. Barry answered in detail and helped students understand the daily routine as well as some of the scientific aspects of space exploration. One participant commented, "We have had fun not only in actually doing the videoconference but also in preparing and debriefing afterwards. It has jump-started our students' study of astronomy and also increased Larry's interest in doing videoconferencing with other academic projects.

Additionally, we had Senator Edward Kennedy talk with students about the telecommunications bill; Representative Maxine Waters from Watts in Los Angeles talk with students during Black History month about the issues facing disadvantaged youth.

Successful Partnerships

The events are the co-invention of several different types of organizations (educational, cultural, scientific, business, technology) playing different roles as providers and consumers of each others' content, pedagogies, technologies, and intellectual resources. For example, in March 1997, we had schools view a broadcast of a live musical piece never before played over the Internet. Afterwards students talked with the composer and the musicians and submit commentary on the musical piece. Partners with the National School Network in this live performance include the broadcast station WGBH, AudioNet, web-based ICHAT, RealAudio, and The New England Conservatory of Music as well as schools around the country. (look up text from chat for real questions)

Local Model

Underlying the "Exchange" is the idea that it can be replicated within a local community or school using the tools and techniques developed in the National School Network "Exchange". The key to the success for a local "Exchange" is a community partnership program which proactively targets and engages representative corporations, small businesses and institutions from a local community or region and its schools, to work on specific community-based, school reform, and local infrastructure issues. The "Exchange", as the online environment supporting this advanced community partnership with schools, will be the mechanism through which these partners provide and exchange local data, seek technical resources and know how, and make visible to other communities the work they do together. We ask community institutions to provide program and information assets online. A particular focus for corporate partnerships are those companies that can provide technical and applications expertise to the burgeoning technical expertise among students and teachers within schools. An example of a local partnership entails the MediaOne Cable Company bringing unprecedented technical expertise in collaboration with BBN, Corporation's Learning Systems and Technology Department in a pilot "cyber-mentoring" project in the Watertown Schools. During the summer MediaOne employees are using BBN's Mentor Center™ software available on the "Exchange" to mentor teachers and by fall, mentoring will be extended to include students and classroom assignments. In the fall, BBN will support the ongoing collaboration in Watertown with an Watertown Teachers Exchange which will include ongoing reports of school-community collaboration, host "Back to School" events where students talk with experts to explore ideas about the classroom of the future, and will provide parents with information and guides to help them understand what their children are learning.

Acknowledgements
Thanks to Elaine Crowder, Beverly Hunter, Hester Fuller, Gail Pierson, Ron Schachter, Valerie Smith for their work on the Exchange and the events. This project was funded by the National Science Foundation, Grant Number RED-9454769

Mallard: A Java Enhanced Learning Environment

Charles R. Graham, Michael L. Swafford, Donna J. Brown
Department of Electrical and Computer Engineering
University of Illinois at Urbana, U.S.A.
crgraham@uiuc.edu, swafford@uiuc.edu, djb@uiuc.edu

I. Introduction

The past several years have ushered in a new era in computing, a consequence of unprecedented growth and change in the World-Wide Web (WWW). One of the most exciting developments is the integration of the Java programming language into WWW browsers. Benefits of using the WWW to enhance education were recognized by the developers of Mallard, a WWW-based educational system developed at the University of Illinois at Urbana-Champaign [Swafford & Brown, 96a][Swafford et al., 96b]. Mallard has been used by thousands of students in a dozen different courses at the University of Illinois (see Mallard homepage at http://www.cen.uiuc.edu/Mallard/). This paper describes some of the ways that Java has been integrated into the Mallard learning environment in order to enhance the educational experience of students at the University of Illinois. The original version of Mallard relied solely on HTML forms and CGI scripts to correct and grade student answers, but the subsequent use of Java has allowed the creation of higher quality lessons and exercises. Specifically, using Java has allowed us to add client-based simulations, visualization tools, and a high degree of interactivity to Mallard homeworks and lessons.

II. Improved Visualization using Java

One of the most fundamental needs in asynchronous learning involves the creation of annotated diagrams. If annotations are static (as in traditional textbooks), then simple inline images are sufficient. However, if the annotations, such as parameter values in homework problems, change over time or are randomly generated for each student, then the use of Java to dynamically annotate the diagrams is practical and provides visually appealing diagrams. In our introductory electrical engineering course, we draw resistive circuits to teach students Kirchhoff's Voltage and Current Laws. In order to present each student with a different circuit to analyze, the resistor and source values are randomly generated. Without the use of Java there are basically two ways to present the annotated figures. First, a new image for each possible set of circuit values could be generated. This method is impractical because it wastes server resources and it is also very labor intensive to develop a large number of images. The second method of presenting an annotated figure is to give all of the students the same circuit diagram labeled with symbolic variables (see Figure 1). The randomized resistor and source values represented by the symbolic variables could then be listed next to the figure. Although this is sufficient to allow students to work the problems, it is far less convenient (especially when dealing with large circuits). It would be ideal to provide circuit diagrams for the students that have actual resistance and source values labeled on the circuit. To solve this problem we developed a simple Java applet to put textual annotations on top of images (see Figure 2).

In Mallard we use this applet to display images, such as circuit diagrams, with randomized values. This simple applet helps illustrate the power of Java; it is not only possible, but also practical to create WWW-based course material with extended capabilities (such as using randomized values) without compromising the quality of presentation found in traditional textbooks.

III. Increased Interactivity using Java

One of the ways in which Java has helped to transform the way educational materials are presented on the WWW is through increasing student interactivity with the learning materials. In the traditional textbook approach to doing homework, it is impossible for a student to interact with the homework, but, in WWW based learning environments, interaction between the student and the learning materials is essential. In the original version of Mallard [Swafford & Brown, 96a], this interaction was handled using HTML forms and CGI scripts and Mallard could grade any problem with an objective solution that could be represented using text.

Although many problems have solutions that can be represented textually, not all are easily represented in a textual format. Also, the textual solution for these kinds of problems may not be as intuitive to the students or teach the concepts as effectively as more graphical formats. Newer versions of Mallard use Java and JavaScript in conjunction with HTML forms in order to maintain a high level of interactivity with the student without compromising the visual quality of problems. Mallard does this by using Java applets to implement a graphical front-end which interacts with the students, translates the students' solutions to a textual form, and sends the converted solutions to the Mallard engine via Netscape's LiveConnect. Mallard is then able to grade the solutions and provide the students with important feedback as it would with a non-Java enhanced problem. This section explains and gives some examples of how Java is currently being used by the Mallard framework to create an interface for problems that are inherently non-textual.

III-A. Timing Diagram Applet

One of the courses using Mallard is an introductory Computer Engineering course. An important part of this course is teaching students to understand and draw Boolean timing diagrams. Traditionally, this skill is tested by requiring students to draw circuit timing diagrams by hand and submitting them to be graded by an instructor. This type of graphically intensive problem does not generally translate well to HTML forms. By using Java, we have created a graphical front-end that allows students to "draw" the timing diagram in an intuitive manner and either have it corrected by the applet or submitted to the Mallard server to be graded (see Figure 3).

III-B. Piece-wise Linear Graphing Applet

Another common type of problem requires students to represent computed data by drawing a two-dimensional graph. The act of drawing a graphical solution can help students to gain insight because they are able to visually see the correlation between different variants in the problem. Because HTML forms provide a text based medium for transferring information through the internet, they are not ideal for use in representing graphical solutions to problems. If HTML forms alone were used to evaluate solutions, a student would have to supply a textual representation of the answer (such as a list of points or equations and ranges) to the server for grading. This would be both tedious and less insightful than actually drawing a graph.

Integrating the use of Java into Mallard has allowed students to actually draw graphical solutions and have them graded. For example, in a freshman electrical and computer engineering course, students are taught to use graphical methods to determine the gain in transistors, to graph I/V characteristics for circuits, and to graph output waveforms for capacitive and inductive circuits. An applet has been developed which allows the students to draw a piece-wise linear graph in their browser windows. When the student has completed a graph, the applet uses LiveConnect to send the essential information to Mallard for grading.

IV. Java Simulation Tools

Another area in which Java has helped to enhance student learning is through the development of client-based simulation tools. As part of the introductory Electrical and Computer Engineering course, students learn basic assembly code programming skills. In order to give the students first hand experience in writing assembly code for a microprocessor, a simulator applet was developed using the Java programming language. The microprocessor simulator called the Knight2000 implements a simple 16 instruction microprocessor with the corresponding memory subsystem and I/O ports [Graham 97a]. The simulator allows the students to write and execute programs online. It also interfaces with Mallard so that the students can have their assembly code programs graded online [Graham & Trick, 97b]. Simulators such as the Knight2000 and others can be powerful teaching tools because they interact with the students and simplify tedious grading tasks.

V. Conclusion

The development of the Java programming language can be a huge help in overcoming some of the practical barriers to providing a seamless environment in which students can learn via the WWW. By using Java we have been able to make great improvements in Mallard. We have used a visualization applet to display diagrams with more flexibility than a traditional textbook or HTML page, and we have also developed applets that allow students to answer questions in an intuitive and graphical manner. Finally, the development of the Knight2000 simulator applet allows a level of immersion in the course material that is not possible using a traditional textbook. While these accomplishments are great enhancements to WWW-based learning, the technologies of Java and the WWW are both relatively young, and we must be ready to make better utilization of them. The integration of Java and the WWW certainly has had and will continue to have a profound effect on WWW-based education.

Works Cited

[Swafford and Brown, 96a] Swafford, M.L., & Brown, D.J. (1996). "Mallard: Asynchronous Learning on the World-Wide Web," Proceedings of the ASEE 96 Conference, Washington, DC, June 23-26, 1996.

[Swafford et al., 96b] Swafford, M.L., Graham, C.R., Brown, D.J., and Trick, T.N. (1996). "Mallard: Asynchronous Learning in Two Engineering Courses," Proceedings of the Frontiers in Education 96 Conference, Salt Lake City, UT, November 6-9, 1996.

[Graham 97a] Graham, C.R. (1997). "Enhancing Learning Environments Using the Java Programming Language," Masters Thesis, University of Illinois, Urbana, Illinois, Department of Electrical and Computer Engineering, August 1997.

[Graham & Trick, 97b] Graham, C.R. and Trick, T.N. (1997). "An Innovative Approach to Asynchronous Learning using Mallard: Application of Java Applets in a Freshman Course," To be published in Proceedings of the ASEE/IEEE Frontiers in Education 97 Conference, Pittsburgh, Pennsylvania, November 5-8, 1997.

Comparison of Two Browser Interfaces: Impact of Browser Migration to On-Going Design

Lee Ann Grim, AT&T, Operations Technology Center
Atlanta, Georgia, USA   lagrim@ems.att.com

The purpose of this paper is to describe a recent user interface work effort in a large telecommunications company. A team of Human Factors Engineers (hereafter called the Standards Team) were asked to develop a common-look-and-feel standards document for web-based user interfaces using the Microsoft Internet Explorer (IE) browser. Many of the team's user interface decisions were based on IE defaults and the look and feel of the IE browser itself. Near the end of the work effort, the governing architecture entity directed the Standards Team to also support the use of the Netscape Navigator browser, which has a different look and feel from IE. The Standards Team did not have time to make changes to the first release of the Standards Document. However, very soon, the author of this paper will be reevaluating the current decisions to determine if Netscape will change the look and feel enough to warrant revisions to the current standards. This paper describes some of the decisions made by the Standards Team and a first guess by the author on how the decisions may need to be modified because of the inclusion of Netscape. A 'Lessons Learned' section has also been included in the hopes of making similar work efforts easier for other user interface design teams.

Because of time and resource constraints, the Standards Team limited their scope of decision-making to desktop personal computers with 17 inch monitors operating in the Windows environment. The first release of the Standards document included the categories of color, fonts, browser defaults, navigation, hypertext links, graphical links, and frames.

Fonts:

Even though users can customize font types and sizes, the Standards Team decided to recommend specific fonts and resolutions after receiving many requests for that information. For IE, the team selected Arial Medium with a resolution of 1024 x 768, or Arial Small with a resolution of 800 x 600. Netscape classifies fonts differently and provides specific numerical point sizes instead of categorical sizes like IE (e.g. small, medium, large) The team performed a visual comparison between IE fonts and Netscape fonts, and recommended Arial 12 in the 1024 x 768 resolution, and Arial 10 in the 800 x 600 resolution for Netscape. These font types and sizes were chosen because they corresponded to the fonts recommended in an existing company GUI standards document.

Colors:

As with fonts, users can customize colors. The team decided to recommend a standard after receiving many requests for help in this area. The team selected light gray (hexidecimal "EFEFEF") for general backgrounds and black for general text. These colors corresponded to the colors recommended in the existing company GUI standards document. The team learned that the appearance of some colors can be different depending on the browser that is used. Colors are also affected by hardware equipment and operating systems. In some cases, the light gray appeared to be almost off-white, in other cases it appeared light yellow. The color decision may need to be modified to a color that is browser independent.

Browser Defaults:

The team recommended that the defaults built into the IE browser for color and identification settings for normal (non-traveled) links and traveled links not be changed. The default colors are blue for normal links and purple for traveled links. The links are identified with the use of an underline. The default settings within Netcape are the same as IE, so additional investigation was not needed in this area.

Frames:

The team recommended the use of frames in specific templates. Both IE and Netscape allow scrollable frames to be displayed without a visible scrollbar/elevator. Only a thin border appears, and the frame can still be scrolled. This gives the page a clean, less-cluttered look. Both browsers allow users to move the borders of the frame. When the wordwrap HTML tag is specified, both browsers will attempt to word-wrap the contents of the frame when the frame is made smaller. When the frame becomes too small to allow the word-wrap of the contents, the frame will overlay the contents. The standards for frames will not need to be modified because of the new browser.

Navigation:

The Standards Team investigated the use of the Back, Forward and Home buttons on the button bar. The team recommended that applications provide their own internal application navigation where needed. The Back and Forward buttons should not be used because the navigation results are unpredictable, especially in frames. The application's internal navigation buttons should be labeled with the name of the page to which they are taking the user. This standard will not have to be re-visited for Netscape.

Page Titles:

The team recommended that an appropriate title be provided in the browser-provided title bar, for each page in the application. This information is altered by both browsers. IE adds the words "Microsoft Internet Explorer" after the title. Netscape inserts the word "Netscape" before the title and encloses the title in brackets [ ]. The browser names are not included in the bookmark/favorite title when a bookmark/favorite is added. This standard will probably not be re-visited.

Lessons Learned:

The Standards work will continue throughout 1997 as time and headcount allow. Here are some Lessons Learned which may benefit other teams who are doing web-based design and development work:

- If the team members are located in multiple geographic sites, each site should have access to a PC with web access; this will ensure that during team meetings the team members are looking at the same pages at the same time.
- All team members' PCs should be set to identical resolution, font size, colors etc.; this will ensure that the team is looking at identical pages, widgets, documents, etc. when design decisions are made.
- If comparisons need to be made between sample designs (e.g. different layouts for horizontal text links), it is helpful to have the samples available for viewing before the meeting at which decisions will be made. This gives team members the opportunity to look at the samples on their own time and to be better prepared to discuss the options at the team meeting.
- All user interface design options should be usability-tested as much as possible.

Acknowledgments

The author of the paper wishes to acknowledge the assistance and support of the Standards Team members and the support managers, John Eddy, David Lu, Max Schoeffler, Pete Spring, Bonnie Strand,and Tom Weber.

On-line Access to the 17th Century Literature

H. Haddouti, S. Wiesener
Bavarian Research Center for Knowledge-Based Systems (FORWISS), Germany
email: [haddouti, wiesener]@forwiss.tu-muenchen.de

Introduction

The development of a high technological network infrastructure and the rapid proliferation of information resources make digital libraries one of the important challenges of computer science. Besides making information retrieval and delivery more comfortable digital libraries can support preservation also. They can provide online access to historical and cultural documents whose existence is endangered due to physical decay.

Over the last three years FORWISS (Bavarian Research Center for Knowledge-Based Systems) has been involved in a project to collect an initial core of all prints of the 17th Century published in the German-speaking area in digital form - VD17. This core has been built up in collaboration with the German libraries of Berlin, Dresden, Gotha, Halle, Munich and Wolfenbüttel. Other national and international libraries are showing interest.

In the long-term project VD17[1], the distributed digital library system OMNIS [Bayer 93], [Bayer, Vogel, Wiesener 94] has been developed to manage more than 300,000 catalog entries and 1,2 million pixel images of scanned key-pages. High-speed networking and the potential of Internet-based technologies - such as the World Wide Web (WWW) - enhance OMNIS to provide fast and world-wide access to the historical prints for users from various areas.

VD17 is based on OMNIS, which is arranged in a distributed client/server architecture. The atomic unit for the archiving and retrieval process is the "document" which may correspond to several catalog entries in VD17 and provides information in different forms: a lot of attributes, a fulltext form and a sequence of images.

In the following sections the workflow of the VD17 project will be described in more detail.

Catalog Management

A major issue of the VD17 project is the integration of a legacy catalog registration system with the digital library system OMNIS. In more detail, the structure of catalog data as it is provided by the legacy registration system is organized in large sets of trees: Librarians organize these data using the so-called MAB-Format (Machine Exchange Format for Libraries), which is strictly hierarchically organized and serves data exchange purposes. These hierarchies allow the librarians an efficient and consistent catalog management. However, browsing and retrieval in this hierarchical structure is expendable and inefficient.

---

[1] VD17 is funded by DFG (Deutsche Forschungsgemeinschaft)

Therefore, the catalog data are managed in two logically independent databases redundantly. On the one hand, the catalog data is stored in a relational data model, which describes the hierarchical structure of catalog trees, their identity, their relationships to each other and their attributes. On the other hand, the data organization of OMNIS is based on flat documents in which all relevant catalog entry information of one print is accumulated, formatted and supplies a fulltext form and structure field entries of a single OMNIS document.

In addition, the cooperative registration (in Berlin, Dresden, Gotha, Halle, Munich and Wolfenbüttel) requires a synchronization concept in order to avoid inconsistent registrations, e.g. duplicate entries.

Image Management

It is estimated that for each title (catalog entry) four relevant key-pages have to be scanned and archived. For 300,000 catalog entries approximately 1,2 million pixel images will be scanned and stored in distributed image databases. As storage media hard disks and CD-ROM jukeboxes are used.

Thus, the distributed image databases allow decentralized image management [Dörr, haddouti and Wiesener 96]. To satisfy the issues of image archiving (high quality, low storage costs and quick network access) images are scanned with 1-bit color depth (black-white) in a resolution of 300 dpi, and compressed with loss-free TIFF G4. The average size of compressed images is about 65 KB. Registration IDs are dispensed for catalog data and scanned key-pages to ensure a mapping of a catalog data entry to its images. The image databases contain images, stored in BLOBs (Binary Large OBjects) [Meyer-Wegener 91], and further image attributes such as registration ID, resolution, size, used compression, format, etc. First, scanned key-pages are stored into image databases on hard disk. If such a database reaches the maximum size of one CD, it will be written to a CD which represents an independent image database managed by a Jukebox. To make this image database available to the catalog database and finally accessible to the users, it has to be announced to the corresponding catalog database.

WWW Gateway

Beside a special OMNIS-client a WWW-gateway [Clausnitzer, Vogel and Wiesener 95] was developed to provide unlimited and platform-independent access to highly valuable, historical heritage. The OMNIS-client provides full text and structure fields queries. Wildcards, boolean operators and phrases are supported. Structure fields serve to use a traditional retrieval, e.g. giving author name, title, etc.

Processing the query *jugend & (laster% | moral)* in the pilot project Oettingen-Wallerstein will deliver all documents (bibliographic data and pixel images) that contain the word *jugend* and either a word beginning with *laster* or the word *moral*. One of the retrieved pixel images is shown in Figure 1.

To make the digital library system OMNIS available to WWW clients an OMNIS-WWW gateway was developed (Figure 1). The implementation is based on the CGI (Common Gateway Interface) definition. CGI defines how WWW server and script (CGI program) communicate. This feature allows the creation of dynamic documents on the fly. Incoming client requests to OMNIS will be received by a HTTP server [Berners-Lee 93] which opens the communication channel to the gateway. The CGI program reads a global configuration file which contains OMNIS server addresses and how the results will be handled. It includes also references to HTML masks. These masks consist of fixed HTML fragments which are filled by the gateway application. These fragments contain also hidden database queries and other commands.

Thus, the CGI application analyses the user's input, parses it and translates it to the database queries which will be passed to the OMNIS server for execution. The query results are sent back to the CGI program which transforms them into a HTML page and passes it back to the HTTP server again. Finally, the HTTP server returns it to the browser for display.

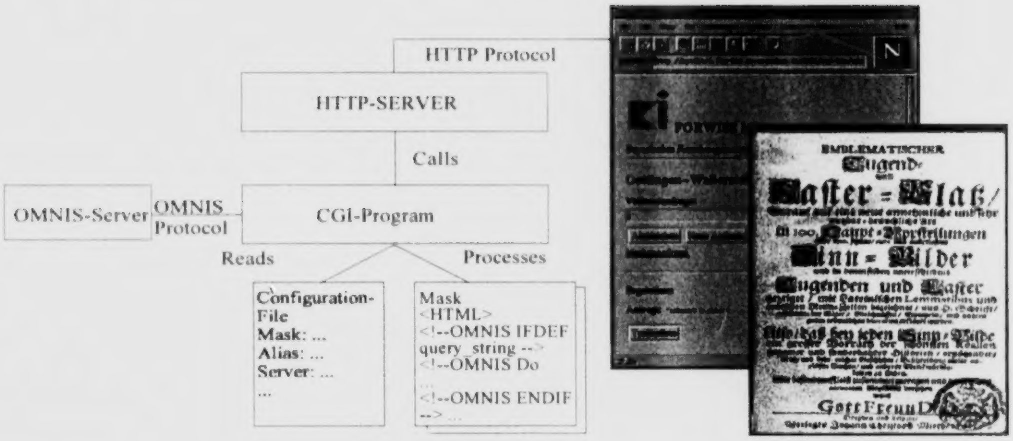

HTTP Protocol
HTTP-SERVER
Calls
OMNIS-Server
OMNIS Protocol
CGI-Program
Reads
Processes
Configuration-File
Mask: ...
Alias: ...
Server: ...
...
Mask
<HTML>
<!--OMNIS IFDEF query_string -->
<!--OMNIS Do
...
<!--OMNIS ENDIF
--> ...
EMBLEMATISCHER
Tugend- und Laster-Blat/
Sinn-Bilder
Tugenden und Laster
Gott Freund

Figure 1: WWW-Gateway

More information is available at:

http:/www.forwiss.tu-muenchen.de/~vd17 (The VD17 project)

http://www.forwiss.tu-muenchen.de/~oewal (The pilot project Oettingen-Wallerstein)

References

[Bayer 93] R. Bayer: OMNIS/Myriad: Electronic Administration and Publication of Multimedia Documents, Informatik Wirtschaft und Gesellschaft, 23. GI-Jahrestagung, Springer, Dresden, 1993

[Bayer, Vogel, Wiesener 94] R. Bayer, P. Vogel, S. Wiesener: OMNIS/Myriad Document Retrieval and Its Database Requirements, DEXA 94 (Database and Expert Systems Applications), Proceedings, Springer Verlag, Berlin, 1994

[Berners-Lee 93] T. Berners-Lee, R. Fielding, H. Nielson: Hypertext Transfert Protocol - HTTP/1.0, 1993. http://www.w3.org/hypertext/WWW/Protocols/HTTP1.0/draft-ietf-http-spec.html

[Clausnitzer, Vogel and Wiesener] A. Clausnitzer, P. Vogel, S. Wiesener: A WWW Interface to the OMNIS/Myriad Literature Retrieval Engine, Third International World-Wide Web Conference, to be published in COMPUTER NETWORKS AND ISDN SYSTEMS, Elsevier/North Holland, Amsterdam, 1995

[Dörr, haddouti and Wiesener] M. Dörr, H. Haddouti, S.Wiesener: Das 17. Jahrhundert im Netz, DFN-Mitteilungen, No.41, June 96, DFN-Verein, Berlin, 1996

[Meyer-Wegener 91] K. Meyer-Wegener: Multimedia Databases, Teubner, Stuttgart, 1991

BEST COPY AVAILABLE

Construction of Consulting Server

Kenichi Hagiwara
Fuji Electric Co., Ltd., Japan
hagiwara@fujielectric.co.jp

Introduction

Users of equipment and computer system need information on usage, trouble countermeasures or maintenance of the products. Products makers provide manuals and helpdesk through telephone or FAX for this information. Today according to internet expansion such online services begin to work on the network that users can get necessary information by replying to a series of questions from the service system.

Many of information services on internet today ask a question to user, get an answer one by one and decide the next question. This method causes the following problems. Users can not look over the retrieval process, and sometimes feel frustration. It is difficult to combine a set of user input and to reach appropriate guidance in the consulting process. In addition this method increases network transaction.

In this paper "consulting server" is proposed whose goal is to provide users with consulting service which is more flexible and efficient on the network than today's method.

Configuration of Consulting Server

Consulting server consists of the following elements:

(1)Web Server with Transaction Process Monitor
This is general type Web server. In order to manage transaction process TP monitor may be used. Web server interacts with Web clients in HTTP protocol. The server shows questionnaire forms in HTML to the clients. Users input to the forms and the Web server gets the replies. HTML pages often include multimedia applets. Video movie applets help users to understand how to check required findings for example.

(2)Database/Rulebase System
In general database stores consulting documents or case files which are records of past consultation. Rulebase stores rules which link users input to appropriate guidance. Information in the database/rulebase is retrieved by a database/rulebase engine in the process of consultation.

(3)Consulting Memory for each client
Consulting Memory stores the process of database/rulebase retrieval which is written by the database/rulebase engine. Such information as users input, list of retrieved data or intermediate hypotheses which are inferred from user input is recorded in this memory.

(4)Consulting Engine
Consulting Engine analyzes the content of the consulting memory to select a set of the succeeding questions and generate a questionnaire HTML page. This page is passed to the Web server to be displayed in the next transaction. This engine also receives users input through the Web server and sets them in the consulting memory. After that the engine invokes the database/rulebase system to retrieve in the new situation and update the content of the consulting memory.

Structure of Consulting Memory

Consulting process is recorded as a list of "consulting nodes" in the consulting memory. A consulting node has the fields shown in [Fig.1]. Each node has its unique "item name". It can have a "explanation URL" which is a link to explanation of the item such as a video movie file or a HTML page. Each node has "value" and its "value type". Logical type, arithmetic type and URL type nodes have true/false value, integer/floating value and URL which specifies the value file respectively.

The node can have "operator" and its "operands". The number of the operands is defined by the operator. Nodes in the consulting memory are linked to each other with operand fields according to the consulting process. [Fig.1] shows an example of the consulting memory which records the following process: A user input two findings Fnd1 and Fnd2. Thereafter rulebase engine inferred that an intermediate hypothesis Int1 was satisfied and a hypothesis Hyp1 could be established if finding Fnd3 was satisfied.

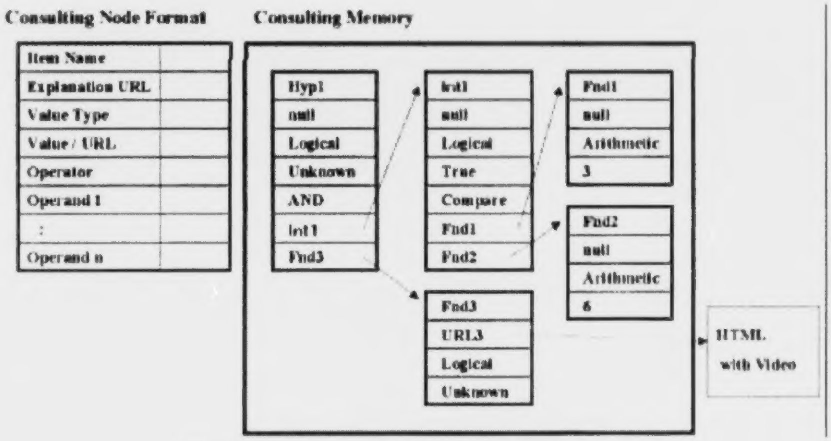

Consulting Node Format
Consulting Memory
Item Name
Explanation URL
Value Type
Value / URL
Operator
Operand 1
:
Operand n
Hyp1
null
Logical
Unknown
AND
Int1
Fnd3
Int1
null
Logical
True
Compare
Fnd1
Fnd2
Fnd1
null
Arithmetic
3
Fnd2
null
Arithmetic
6
Fnd3
URL3
Logical
Unknown
HTML
with Video

Figure 1: Structure of the Consulting Memory

Control of Consulting Process

By means of analyzing the consulting memory, the consulting engine can get useful information and control the consulting process efficiently. It can restrict useless questions to users previously. If the intermediate hypothesis Int1 had value false for example in [Fig.1], the finding Fnd3 is not to be included in the succeeding questionnaire page because it has no effect on Hyp1 establishment.

It can also combine a set of information and identify critical information which influences the retrieval result. Suppose that a user set the finding Fnd3's value unknown in [Fig.1] for example. The consulting engine will generate a questionnaire page which includes Fnd3 again, because the consulting engine finds out that the finding Fnd3 is the critical node which influences the hypothesis Hyp1's establishment in combination with the intermediate hypothesis Int1.

Users can understand the problem solving process, because they can look over the related questions in the display. At the same time network transaction is reduced.

Summary

Consulting server is proposed in which the consulting memory mediates between the Web server and the consulting database/rulebase and the consulting engine controls the consulting process by referring to the consulting memory.

I think it convenient that the structure and the format of information in the consulting memory are

BEST COPY AVAILABLE

open. This leads the consulting server to be open to several existing database/rulebase systems. And the application interface of the consulting engine should be open in order to be used in the widespread Web servers.

Student Evaluation of a Web Based Course

Dwayne Harapnuik, MLS
Division of Technology in Education
3-104 Education North
University of Alberta
Edmonton, Alberta
Canada T6G 2G5
E-Mail: dwayne@oanet.com

T. Craig Montgomerie, Ph.D.
Division of Technology in Education
3-104 Education North
University of Alberta
Edmonton, Alberta
Canada T6G 2G5
E-Mail: Craig.Montgomerie@ualberta.ca

Introduction

This paper discusses student evaluations of a credit course: The Internet: Communicating, Accessing & Providing Information which, is delivered completely over the Internet. Since May of 1996, approximately 300 students have completed the online course. Upon completion students are asked to provide a detailed evaluation of their learning experience. This paper provides a brief review of the structure of the course, a discussion of the development of the survey instrument, and the data collection methods. Examples of the kinds of data collected and a summarized statistical analysis will also be provided. Student generated comments from the evaluations will be summarized and an examination of their learning experience will be highlighted. From this summary, a list of conclusions will drawn as the basis of a set of recommendations for the development of this particular course and future courses.

Course Structure

The course is designed to follow good andragogical (adult learning) principles; particularly that the user should be in control of their own learning (content, pacing, and sequencing), that alternative methods of learning the same material should be available, and that the subject area for assignments should, if possible, be the student's choice. One of the objectives of this course is to get students accustomed to seeking out their own answers. This fits extremely well with principles of adult learning, and the philosophy of life-long learning. The Internet is changing so rapidly that it is very difficult to predict accurately what a person will need to know or understand a year from now. Internet courses must prepare students to accept the responsibility of learning and help them establish patterns of searching out new information on their own. The course content happens to fit perfectly with the educational philosophy and principles which we espouse and the subject matter of the course.

Survey Instrument

Since the course is delivered entirely online, the survey instrument was made available in the same fashion. Web-based forms allow for simplified collection of data, and data from the HMTL can be moved directly into a database or into one of many different statistical analysis programs. The survey is comprised of two main sections. Section A includes: Universal Student Ratings of Instruction, which are instructor evaluations that are standardized within our university. Section B includes: Course Evaluation, Personal Information, and Open Ended questions.

Section A was implemented to replace the traditional instructor evaluation forms that most (if not all) academic institutions use to evaluate instructor performance. The results of the seven questions and additional comment section are forwarded to the department chair.

The Course Evaluation component of section B is comprised of twenty closed ended question that ask students to Strongly Disagree, Disagree, Neutral, Agree, or Strongly Agree. The actual questions range in scope from asking about instructor approachability, availability, and support to queries on the amount of work, type of assignments and degree of difficulty. The HTML form requires that students check off only one response.

The Personal Information component includes eight questions designed to determine the student's educational background, faculty or program, status and future plans. In addition, students are asked about their computing background and the type of computer equipment they own or purchased to complete the course.

The final component of section B includes five open ended questions which asked what the students liked most or least about the course, what changes they would like to see implemented, and if they would be interested in participating in additional courses delivered via the Internet.

Student Evaluations

The on-line environment allows students to be very candid. Shy people often find courage to say things they may not have said in a F2F (Face to Face) situation. This allows for some exceptional feedback. One of the most common responses to the course in general can be summed up by a paraphrase of one student's e-mail message:

"Taking this course was one of the worst things that I have done and one of the best things that I have done. It is the worst because I now spend all my spare time on the Net and it is the best thing because I have learned about what is out there and can now not only access all that information, but I can also contribute."

Approximately 40% of the students in the 1996 spring and summer session of the course filled out the survey and about another 10% of the class submitted an E-mail evaluation. Similar participation results were found for the 1996 fall session 1997 winter session of the course. The final results for the current session 1997 spring and summer session will be made available in early 1998. Due to flexible start and completion dates, most winter 97 students completed the course during the spring and summer session of 1997 and used a survey mechanism that is part of a newly implemented Course Administration and Database Management system.

The following is a brief summary of the open comments sections from the spring, summer and fall sessions of 1996. A detailed breakdown of this data as well as full details on winter, spring and summer sessions of 1997 will be offered in a subsequent full paper on this topic. Since the course is an official Faculty of Education course it is not surprising to have approximately 50% of the students from Education. What is surprising is that more than 25% of the students are registered as unclassified students. Many people from business or industry have heard of the course and participate in it because they may be responsible for their department or companies Web site.

Students either hated or loved the course (or both!). In general, most students felt that the volume of course work was greater than that for most other courses. Most students also indicated that they were not comfortable with the on-line format and missed the F2F component of typical instruction. Those who enjoyed the course really appreciated the flexibility of working on the course at their own pace and at their own time. In contrast approximately one third of the students stated that they would have liked a fixed schedule and specific assignment dates. Most students found the course conferencing system a useful replacement for F2F interaction but were not satisfied with the actual conferencing software.

One of the most startling revelations was the paradox that most students expressed. Approximately 70% of the students stated that they took the course because it was delivered completely online and offered the greatest amount of flexibility, yet almost 90% of respondents stated that they desired some sort of face to face instruction. The range of F2F instruction requested started at one F2F information session at the beginning of the course to a request for weekly labs.

Recommendations

There is no doubt in our minds that a very effective and efficient instruction can be delivered over the Web and that this course is moving in the right direction. One community of scholars, the North American Web Developers Association, awarded this course "Best Educational Web Site: Single Course" (NAWeb, 1996). Student learning, as evidenced by their comments and their assignments, was similar for students in the on-line version of the course and those in a F2F mode. We have learned a number of things while designing and delivering the course:

1. the instructor should have taught the course previously in a F2F mode in order to design the course effectively
2. use good adult learning principles (learner control of content, pacing & sequencing; alternative methods of learning; self-selected assignments)
3. provide good student-student and student-instructor communications
4. be prepared to deal with students who find the technology "gets in the way of learning"
5. make the course load as quickly as possible
6. take advantage of other people's work (other resources on the Web)
7. allow students to help each other and to discuss things among themselves without feeling the need to "guide" all discussions, but provide fast feedback to students who are encountering problems.
8. expect that the workload will be significantly higher for both the student and instructors than for a F2F course on similar material.
9. be prepared to continually modify your course pages
10. always remember, and remind your associates and students, that there are less cues in computer mediated communications than in F2F communications; what is said with a smile can sound harsh when printed
11. enjoy the experience - it is different, but it is teaching!

Questionnaire Administration Via the WWW: A Validation & Reliability Study for a User Satisfaction Questionnaire

Ben D. Harper, Laura A. Slaughter*, and Kent L. Norman

Department of Psychology, Laboratory for Automation Psychology
College of Library and Information Services (CLIS)*
University of Maryland, College Park, MD 20742 USA

1.0 Introduction

The World Wide Web (WWW) was used to collect data for a reliability and validation assessment of a new version of the Questionnaire for User Interaction Satisfaction(QUIS). The use of the WWW for this experiment provided an appropriate population of users to test this particular type of questionnaire, standardized questionnaire administration to participants, and made data processing virtually effortless. In addition, this method of testing was done at a lower cost and took less time for data collection than a traditional experiment of the same nature. In addition, this experiment revealed the numerous questions that must be considered when using the WWW as a tool for collecting data.

There are three primary issues concerning the use of the WWW for data collection. These are subject characteristics, materials, and administration procedures. Subject characteristics refers to sampling the population of WWW users. Can the WWW be a solution for experimental designs requiring a large sample of computer users? What are the characteristics of the desired population? Materials refers to the use of resources to produce attractive design of layout, dynamically changing questions, automated data processing, multimedia presentations, and elimination of special tasks that would normally require human intervention (such as timing). Procedural issues are focused on how to provide informed consent, debrief participants, deciding on methods for solicitation, and privacy concerns. Aspects from each of these three played some part in the design of this study.

2.0 Materials

The Questionnaire for User Interaction Satisfaction (QUIS) was created to gauge the satisfaction aspect of software usability in a standard, reliable, and valid way. The QUIS 7.0 is an updated and expanded version of the previously validated QUIS 5.5 [Chin et al., 1988]. The QUIS 7.0 is arranged in a hierarchical format and contains: (1) a demographic questionnaire, (2) six scales that measure overall reaction ratings of the system, (3) four measures of specific interface factors: screen factors, terminology and system feedback, learning factors, system capabilities, and (4) optional sections to evaluate specific components of the system: technical manuals and on-line help, on-line tutorials, multimedia, Internet access and software installation. Each item in the questionnaire is rated on a scale from 1 to 9 with positive adjectives anchoring the right end and negative anchoring the left. In addition, "not applicable" is listed as a choice. Users also have the ability to add comments within the questionnaire.

The questionnaire was implemented using standard HTML forms, and its style is very similar to the paper version of the questionnaire. In order to prompt users to consider each question, a response was required for each question. Client-side JavaScript was used to both validate the user's responses and gather them into a consistent and standardized format. The data for each section of the QUIS was time stamped and recorded, however, the data was only sent to the server after the entire questionnaire was completed, guaranteeing the questionnaire integrity.

The on-line questionnaire was made available through the World Wide Web(WWW). The subjects learned of the study through advertisements on WWW directories such as Yahoo, human-factors related mailing lists and newsgroups. The subjects began the questionnaire with two introductory pages, the first explained what the experiment was about and the second gave directions for completing the questionnaire. The subjects were able to quit the questionnaire at any time and progress at their own speed. The subject was not permitted to go to the next page without completing all the questions. After completing the QUIS, a comment page was available to the participants for feedback.

Altogether, eighty-eight participants (61 males and 27 females), voluntarily completed the on-line questionnaire. They ranged in age from 14 to 76. Fifty-seven percent stated they had worked more than six months with the software they were rating. Fifty-eight participants rated a WWW browser of their choice, 14 rated a software product they disliked and 16 rated a software product they liked. A total of 29 different software products were evaluated.

3.0 Results

The overall reliability for the QUIS 7.0 is Cronbach's alpha of 0.95. The mean question scores varied from 4.85 to 8.07 with standard deviations ranging between 1.34 and 2.68. Construct validity was measured by correlating item scores with the 6 concurrent general satisfaction questions validated in previous studies. The mean correlation between each main item and a general satisfaction scale ranged between .49 and .61 (SD .09-.12). This suggests that there is good agreement between the new sections of the QUIS and general satisfaction while not being so derivative as to be redundant.

The reliability of this extension to the QUIS (alpha=0.95) yielded similar results as the previous versions of the QUIS (alpha=0.96 & 0.88) [Chin et al., 1988], and is significantly greater then the minimum reliability suggested by Lewis [Lewis, 1995] (alpha= 0.70). The strong relationship between sub-items and items, and then among items in composite sections suggests that there is a hierarchical structure to the questionnaire.

Demographic data for 89 subjects revealed that 70% of the subjects were male, 82% of them ranged in age from 20-45 years, with a mean age of 33 (s.d.=11) . 62% of respondents had between 1 month and 1 year experience with the product they were evaluating, while 26% had more then a year experience.

4.0 Discussion

Although the demographics of this sample are very similar to those found by other surveys of the

Internet population for the same time-frame, there is no way of knowing how this sample might differ from the total user population. Without convergent demographic measurements of the Internet population from on-line questionnaires and other conventional survey methods we will not be able to determine the agreement between random samples of users and the volunteer samples that can be collected from the Internet. Until this information is available, generalizability of results will be somewhat limited.

Where previous versions forced a separation between the effect of interest (task performance) and the measure of that factor, the Web-enabled QUIS may allow a closer linkage between usability and it's measure. Web-enhanced applications may be linked directly to usability measures, improving fidelity.

Additionally, this study has shown the ease at which large N studies can be conducted without labor intensive and expensive single user testing. In addition, a more age and gender diverse, and possibly more representative, subject sample was available. This starkly contrasts the rather homogenous sample available to many studies.

There are lingering issues associated with using the WWW for experimental administration. Firstly, a WWW sample is not as representative of the general population in some ways, including socioeconomic status, and educational levels. Samples from the WWW may also reflect different international populations. Materials may be presented inconsistently, as the experimenter cannot completely control browser preferences. Furthermore, experiments involving deception are not feasible, given that debriefing will occur only following the experiment and subjects can exit the experiment without proper debriefing. Finally, subjects may be putting their privacy at risk, as experimenters attempt participant tracking or utilize "holes" in the technology.

5.0 References

[Chin et al., 1988]
Chin, J.P., Diehl, V.A., & Norman, K.L.(1988). Development of an instrument measuring user satisfaction of the human-computer interface. in *CHI `88 Conference Proceedings: Human Factors in Computing Systems* (New York, 1988), ACM Press, pp. 213-218.

[Lewis, 1995]
Lewis, James R. IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *Internatioal Journal of Human-Computer Interaction. 7*, 2 (1995), 57-78.

The Design, Development, and Implementation of a Virtual Online Classroom: From a Designer's Point of View

Marilyn J. Heath
College of Education, University of Houston, Houston, Texas, USA
mjheath@ghgcorp.com

New technologies are challenging traditional paradigms of instruction because they support powerful alternative vehicles for teaching and learning. One of the more interesting technologies, from both a technical and social viewpoint, is the World Wide Web (WWW). While most educators are exploring the use of the WWW to supplement traditional, face-to-face coursework, some are investigating online learning in "virtual" classrooms that exist only on the Web. A number of universities now offer complete degree programs "on the web." However, most of the courses available on the web today take a traditional approach that involves specification of clear objectives that all students are to accomplish and an "information delivery" model that assumes students will learn the content selected by the instructor. Very little has been done to develop online courses grounded in an alternative paradigm such as constructivism.

This study was carried out to create a virtual online classroom based on constructivist teaching and learning principles. The topic selected for the course being developed was "An Introduction to Distance Education." A secondary goal was to learn more about the processes and issues related to successfully producing a virtual online classroom. This study was also challenged to focus on the unique strengths and features of the delivery medium and to avoid creating simply an electronic version of a traditional course.

The instructional design model that guided the design and development process was Willis' Recursive, Reflective, Design, and Development (R2D2) design model [Willis 1995] which is a non-traditional model based on a constructivist-interpretivist epistemology and social constructivist learning theory. In traditional linear ID models, design and development may be accomplished in a linear sequence through a series of somewhat independent activities. The output for one activity generally serves as input for the next activity. In contrast, in a non-linear model such as R2D2, design and development involve non-linear and recursive processes with frequent interaction, iteration, and change. It progresses from fuzzy to final version in a non-linear, sometimes chaotic, fashion. R2D2 is also a participatory rather than an "expert" ID model. All stakeholders -- designers, developers, various experts, instructors, and potential end-users were involved in this study. Collaborative work produced the look, feel, and function of the product.

Design and development activities for this online course took place in three different areas: (1) the choice, design, and presentation of instructional content and activities, (2) design and development of a user interface, and (3) the design and navigation of a communications/ conferencing tool for carrying out asynchronous discussion. The final version of the course was made up of two major online components: (1) the course website which provided the "entrance" into the class and contained all of the course management elements, and (2) the Hypergroup Conference area where all of the discussions and postings took place. The Hypergroup Conference center utilized features of a web-board with a graphical interface and a threaded listserv system. Students were able to participate either directly from the conference center or through e-mail using listserv features. The website and conference center components could function independently, from a technical point of view, but together they formed the integrated framework for the course.

Course content was organized around a series of "threads." Each thread was introduced in the course website where it required some interactive postings from the students or various other activities. Discussion questions from each thread (in the website) provided direct links to discussion areas in the Hypergroup Conference center. A unique feature of this course design was that it did not require the presence of continual moderation or facilitation. All class discussion questions and activities were planned and posted to the website before the course began. This approach did not limit the amount of student participation. In fact, this design created more traffic than any other online course I have encountered. This course design also lends itself to the use of more than one instructor.

Because constructivism was the underlying concept to instructional design of this course, students were allowed to work at their own pace through the course threads, entering discussions in any order they chose. The goal was to allow students to determine the sequence of their participation in activities and discussions. While this provided flexibility for those who desired it, some students would have preferred a structured approach where the whole class worked on the same topic at the same time.

Questions Answered

One of the purposes of this project was to create a model and to provide information for others who might want to develop a web-based course. Common questions that potential web-base course developers ask:

"Does it take more time to develop an online course?" Instructors who have developed online courses remark that they think developing an online course is much more time-consuming than a traditional course. I am not able to address this issue because I have no current academic course preparation experience with which to compare it. I know that advance course preparation time is considerable, however, once the course is launched, further course development is not generally required. The instructor can be free to actively participate and enjoy the discussions of the students instead of worrying about preparing the next day's assignment.

"Since everything is prepared and planned in advance, doesn't this eliminate spontaneity for both the instructor and the student?" This did not appear to happen. In fact, the students seemed to have a better sense of the direction of the course since they saw it as a whole from the beginning. It did not become a week-to-week activity, or day-to-activity, with instruction being prepared at the last minute.

"How difficult is it to produce a course like this?" I found that it is not necessarily difficult ... just different. How students process the information, what activities are best, how communication takes place --all of these issues need to be considered, because they are different for an online class.

"What skills are necessary for developing an online course?" I have previously created websites, multimedia courseware, and CBT lessons. So, I have a strong understanding of interactivity and navigation through electronic media. If a person has no previous skills creating a website, or understanding navigation, there may be a steep learning curve and the instructor (or developer) will need to find additional expert help.

Here are several recommendations for future online course developers. (1) First of all, don't underestimate the amount of time that it will take to develop a course, especially if it is the first version. If possible, start several weeks (or months) in advance of the course delivery date. (2) Subject matter and content come first. Know your subject inside and out. Organize content into manageable chunks. (3) Decide what types of knowledge or understandings you want students to acquire and what kinds of interactions you want them to have. Remember, students are not all alike, so you will need a diverse range of learning activities. (4) If you have an online conference area, how will you manage the online discussions -- moderated or unmoderated? This will determine how much online time the instructor will need to spend communicating with students. (5) Once you have decided on content and activities, you can make a better choice of the technology tools you want to use. Before you make a final choice, try to find out what the general skill level of your students might be and if your choice of technology is available to them or can be made available. And most of all, (6) each course should be approached individually. There is no right or wrong, or "best" way.

References

[Willis 1995] Willis, J. (1995, November-December). A recursive, reflective instructional design model based on constructivist-interpretivist theory. Educational Technology, 35(6), 5-23.

The Pleasure Principle Revisited in Virtual Reality.

Fran•oise Herrmann
Independent researcher, USA
fherrmann@igc.apc.org
http://www.wenet.net/~herrmann

Background

Development and mature behavior requires surrender of the principle of pleasure and choice of the reality principle [Bettelheim, 1976]. This means action based on reflection, conscious volition and desire in contrast to immediate need-based gratification. But, as Erikson [Erikson, 1946] has pointed out "the concept of 'reality" itself, while clear in its intended meaning, is highly corruptible in its usage", meaning that individual choice of the reality principle without consideration, in the western world, of a third principle, the social principle, all to often leads to both economic and emotional crisis.

To further complexify the social individual's action plans and to continue Erikson's deconstruction of the notion of reality, one might ask in the high-tech post-industrial society of the 1990s whither reality (?) when it has become virtual in VR (Virtual Reality) or ambiguous in the Internet worlds of MUDS and MOOS [Turkle, 1994]. Do the same principles operate for the turn of the twentieth century, socio-techno, individual as they do for the post world-war II, western individual that Erikson saw conflicted by both economic law and those of the psyche? Similarly, but taking up Bettelheim's distinction, one might ask in the Internet world of light speed activity, accelerated and immediate communicative action, whether the principles of pleasure and reality are really mutually exclusive?

Toggling between both of these takes on reality and pleasure, I identify, in this paper, yet another principle: a principle of creativity. Based on an analysis of smileys and ASCII character drawings, I suggest, using a linguistic Saussurian framework [Saussure, 1974] that this choice of linguistic action lies at the intersection of reality and pleasure in Bettelheim's frame of reference. Whereas in Erikson's line of thought, a constructed principle of creativity would yield: "That which is good is that which enables one to break-free of constraint to sustain social interaction, without prejudice to collective and individual comfort both economic and emotional".

Identification of a problematic situation

To the two prevailing explanations, based on compensation and necessity, of the phenomenon of SMILEYS and ASCII character drawings, also termed emoticons, a third explanation is offered based on creative pleasure. This explanation arises *after* examination of the ways in which the absence of standardized meanings across emoticons functions to confound any compensatory value. And as emoticons continue to abound in explosive and forceful ways when other mediational means (e.g.; graphic and image processing programs) are present and readily available.

Method and analyses

The data and findings reported here are gleaned from a five year (and on-going) ethnographic study of an international on-line community of academic scholars using ListServer technology. As a participant-observer I photo-recorded, initially on disk, and subsequently on tapes, the communicative activity, in the form of posted messages, flowing across my computer screen, along with the use of interviews and a survey, while gradually augmenting my own on-line participation (moving from silent participation to one-on-one side-channeled communication; and finally to collaborative writing with a sub-group, and the public posting my own messages to the group.)

Messages containing the use of SMILEYS and ASCII character drawings were collected. For comparative and translation purposes two small dictionary-like reference books were used, as well as a series of list-type files retrieved by search on the Internet. Using a linguistic Saussurian framework, all SMILEYS and ASCII character drawings were seen as signs where the original signifier-signified relationship had been broken and reconstructed via re-assignment of key-stroke function.

Findings and discussion

1> *Smileys and ASCII art drawings are shifted graphics*. These are signs where the arbitrariness of relationship between signifier (keystroke) and signified (keyboard character) has been exploded and reconstructed in creative ways. Colons, semi-colons and commas become "eyes" , for example ":->"; and slashes become "arms and legs" as in the example of the dancing figure [Fig.1]:

```
\ /
 \0/
 V
 |
 /\
/ \
```

Fig 1.

2> *The arbitrariness of the new signifier-signified relationship in the form of a shifted* graphic *is far from standardized.* This is to say that what shifted graphics mean *varies* from one source to another and according to the users' intentions. In Saussurian terms, whereas the arbitrariness of the keyboard character (i.e.; ";" is a semi-colon) is well established, the reconstructed relationship between signifier and signified found in the use of SMILEYS and ASCII character drawings is far from settled. For example, two dictionary definitions of the emoticon ":-]" yield different definitions: "The sarcastic smiley. You are your own worst enemy e.g.; I must have asked people to flame me :-]" [Sanderson & Dougherty, 1993]; and "Classic smiley. Smiley blockhead" [Godin, 1993]. Further, in juxtaposition to usage data, more meanings for the ":-]" smiley appear, as in the following extracted message sign-off:

":-) <----- Note hopeful expression on smiley while awaiting possible change of status in virtue of present message.
:-] <---- Note increased sense of security now expressed in face of smiley
<---- note absence of smiley."

3> *In the presence of alternative mediational means, emoticons continue to abound in potent and explosive ways.* With the advent of the World Wide Web and its capacity to easily process sophisticated motion and still graphics, both Smileys and ASCII art drawings continue to exist in text, and as art galleries. Smiley files retrieved on-line run thousands of lines each listing the shifted graphic and its multiple meanings, for example:

"%-) : after staring at the terminal for 36 hours; broken glasses; cross-eyed; drunk with laughter. "

The most important pertinent consequence of free-floating emoticon meanings is that it confounds compensation functions. Rather than reducing uncertainty, emoticons provide additional dimensions of interpretive space, at least as open as that which is available to the one-reconstructing the signifier-signified relationship (i.e.; the emoticon creator).

Continued use of emoticons when alternative mediational means of expression are available questions exclusive constraints of necessity while suggesting that pleasure is involved. The pleasure that is invoked in discovering new signs; in playing with the arbitrariness of the sign and pushing its limits; and in alignment with archaic explorations of language use. In sum, the pleasure of engaging in creative activity.

Conclusion

Born perhaps out of a desire to compensate and perhaps from the necessities of the keyboard, it is suggested here that emoticons are *also* born of the sheer joy of discovering and expressing in new ways. At the intersection of reality and pleasure, this creative activity enables one to break free of constraints to sustain and enhance communication with a serious playfulness deeply rooted in our early language learning activity.

References

[Bettelheim, 1976] Bettelheim, B. (1976). *The uses of the enchantment: The meaning and importance of fairy tales.* New York, NY: Vintage Books Edition, April 1989.

[Godin, 1993] Godin, S. (1993). *The smiley dictionary.* Berkeley, CA: Peachpit Press.

[Erikson, 1946] Erikson, E.H. (1946). Ego development and historical change. In E. H. Erikson, *Identify and the Life Cycle*, Chapter 1, (pp.17-50). New York, NY: W.W.Norton & Company, Inc. 1980 Edition.
[Sanderson & Dougherty, 1993] Sanderson, D. & D. Dougherty (1993). *Smileys.* Sebastopol, CA: O'Reilly & Assoc.
[Saussure, 1974] Saussure, F. (1974). *Course in general linguistics* . London,UK: Fontana.
[Turkle, 1994] Turkle, S. (1994). Constructions and reconstruction of the self in virtual reality: Playing in the MUDs. *Mind, Culture and Activity* , 1(3), 158-167.

Herbal-T Internet Desktop

Israel Hilerio, i2 Technologies, USA, israel_hilerio@i2.com
Weidong Chen, Engineering & Applied Sciences, Southern Methodist University, USA, wchen@seas.smu.edu

Introduction

Web browsers are the main interface to access World Wide Web (WWW) applications. The clients operating system is no longer a consideration when deploying WWW applications. The WWW has opened the doors to a confusing desktop metaphor where applications are contained and accessed by web documents. Customizing these environments requires scripting language and HTML knowledge. The Herbal-T desktop provides a customizable framework for downloading and building relationships between Java Applets and Applications without HTML intervention. This customizable framework provides a messaging layer that allows Applets and Applications to communicate, coordinate, and share information with each other, independent of server intervention.

This paper discusses the development of an Internet Desktop that combines the application accessibility of a web browser with the customization capabilities of a desktop. Current Internet desktop approaches attempt to extend the web browser to a multi-functional desktop environment (e.g. Netscape). Other approaches attempt to extend the desktop by incorporating browsing capabilities (e.g. Microsoft). Both these approaches use scripting languages and HTML as the bases for their integration. The approach outlined by this paper borrows from these two initiatives and extends their functionality by creating an environment where Applets and Applications can be accessed directly without being contained inside an HTML document. This desktop facilitates the dynamic creation of relationships between downloaded components in the form of Applets and Java Applications without using scripting languages. Relationships can be defined at the class level or instance level.

HTML forces web applications to coexist within the boundaries of a browser. This approach sacrifices local desktop integration and ease of use. Currently, HTML and scripting languages (i.e. JavaScript and VBScript) statically bind relationships between web applications/components. If a user wanted to change the relationships between web components (i.e. Java Applets and ActiveX) he would have to modify the web page or change the scripts. The alternative is an environment that facilitates the dynamic definition of relationship between web components. Through the use of components, users are safeguarded against having to download large monolithic applications that follow the 80/20 rule. This approach leverages a client side communications mechanism that coordinates activities and data between downloaded components in order to reduce network traffic. Client side communications can be accomplished by processing simpler transactions at the client side and providing web components with a piping mechanism between components.

Herbal-T Internet Desktop

The Herbal-T desktop defines an extensible framework in which the functionality of a user environment can be customized through the use of components. The goal is to create an environment where multiple components with specific functionality (e.g. spreadsheet, text editor, and image viewer) can be downloaded over the web and relationships can be dynamically defined between them. These dynamic relationships define the interaction between the downloaded components. Components are downloaded using a URL (http://www.smu.edu/components/textedit.class). Power users will be allowed to customize their environment by selecting components from different web locations to be downloaded into their desktop environment and defining relationship between them. Relationships consist of a <Trigger, Sender, Receiver> tuple that identifies responsibilities between components. Multiple tuples can be associated with the same trigger.

The user selects a trigger action from a downloaded component. This action initiates the tuple transaction. Also, the user selects a sender and a receiver component. The data associated with the sender component will be returned and passed to the receiver component as the input parameter. Relationships can be defined in a single component or separate components. Once the relationships have been defined graphically by the user, the information sharing will be done transparently by the desktop application. Relationships can be defined at the class level before components are instantiated or at the instance level once a component has been created. If a relationship is defined at the class level, all of the instantiated components from that class will share the relationship. If a relationship is defined after the component has been instantiated, the relationship will be unique to that component. Instantiated component can be cloned. Cloned components will share relationships.

Herbal-T defines an extensible Internet desktop environment that provides a classloader responsible for accessing web components through the use of HTTP. Components can consist of Java Applets or Java Applications. Once these components are accessed by the classloader, a windowing framework is provided for each component. In addition to the windowing framework, a functionality outline is defined for each component.
The windowing coordinator manages the windowing framework and the relationships between the native windowing environment and the windowing framework. Part of this functionality allows components to be minimized, maximized, and manipulated across workspaces. Workspaces implement the concept of rooms, where individual environments can be defined in order to allow users to subdivide their work habits. This allows components to be placed in any room based on user preferences.

Another use of the windowing coordinator is its interaction with the broker to manage users' request and access the correct list of functionality available to the broker. This allows the user of a text editor to select the spelling function on the menubar, which sends a request to the broker. The broker communicates back to the windowing coordinator with two available options, Spanish and English. The windowing coordinator takes these two options and creates a pop-up window and displays the information to the text editor user. As far as the user is concerned, all these activities took place at the click of one button without having the user worry about setting any specific flags. The windowing coordinator also allows data files to be matched with web components using an object oriented interface. This allows data files to be accessed by multiple components for execution. If a graphics file is accessed, the corresponding application for viewing the file will be used to display the image. The windowing coordinator defines this relationship.

The functionality outline definition allows the broker or functionality coordinator to manage the passing of information between components without the use of HTML or scripting languages. The main task of the functionality coordinator is to implement transparent component integration. In addition, the functionality coordinator is responsible for coordinating message passing between components, clipboard access, and drag & drop activities. The functionality coordinator is able to negotiate the accessing of information and services between loaded components. The classloader forces components to register their services with the functionality coordinator. The functionality coordinator in turn uses this information to supplement the services of existing components.

The functionality coordinator can also be utilized to create pipelines between components for information transfer. This concept can be illustrated with a graphic illustrator environment. Information related to a specific graphical template may be changed using a text editor component. A link between the text editor component and a specific label component within a graphics template can be specified graphically using the functionality coordinator. This connectivity allows the information typed in the text editor to dynamically be posted on the graphics template.

Conclusion

Herbal-T defines the next generation of Internet desktops that provide a dynamically extensible relationship builder metaphor that requires no HTML or scripting language intervention. This approach allows us to reconcile the integration of web applications into our desktop environment, allowing web applications to be accessed as network applications. Herbal-T allows users to access web applications as applications rather than as documents.

1.Bott, E., & Leonhard W. (1996, November). Windows 97? 98? 99?. **PC Computing,** 154-159.
2. Bott, E., & Leonhard W. (1996, November). Explorer 4.0: New for '97. **PC Computing,** 162-166.
3. Chappell, D. (1996). Understanding ActiveX and OLE (1st ed.) . Washington: Microsoft Press.
4. Cluts, N. W. (1996, July 22). What's New for Internet Explorer 4.0. [On-line]. Available: http://www.microsoft.com/win32dev/ui/ie40.htm
5. Cornell, G. & Horstmann, C. S. (1996). Core Java (1st ed.) . California: SunSoft Press.
6. Ernst, W. (1996). Presenting ActiveX (1st ed.) . Indianapolis: Sams.net Publishing.
7. Gentner, D., Ludolph, F., & Ryan, C. (1996, October 30). Designing The HotJava Views User Interface Environment For a Network Computer. [On-line]. Available: http://java.sun.com/products/hotjava/hjv.white.html
8. Goodman, D. (1996). JavaScript Bible (2nd ed.) . California: IDG Books Worldwide Inc.
9. Lemay, L. & Moncur, M. (1996). Add Interactivity To Your Pages With JavaScript (1st ed.) . Indianapolis: Sams.net Publishing.
10. WebMaster (1996, October 30). Java Beans Reference Paper. [On-line]. Available: http://splash.javasoft.com/beans/WhitePaper.html

Use of Natural Language Technics to Improve Text Retrieval on The Web

C Jacquin, M Liscouet
IUT de Nantes, département informatique, 3 rue maréchal Joffre, 44041 Nantes cedex 01, FRANCE
IRIN, Institut de Recherche en Informatique de Nantes, 2 rue de la houssinière, BP 92 208, 44322 Nantes cedex 03, FRANCE

Email: jacquin@irin.univ-nantes.fr

Introduction

Networks, in particular the Internet, allow large scale retrieval of textual data, but how does one actually find something of interest in this mountain of information and where does a reader find text on a particular subject they are interested in ?.

The Word Wide Web provides tools to search for text by using either simple or composed keywords, which are linked with logical operator. These tools improve the search process, but on the other hand, often retrieve too many references thus resulting in too much text. To improve this search process efficiency further or to allow the user to have an insight into the contents of the text which are to be retrieved by the aforementioned process, terminological tools seems to be the answer. Firstly, in this paper, we outline a terminological extraction tool ANA [Enguehard 1993], developped in our laboratory (IRIN) and then we outline two different uses of this tool in order to improve the document keeping process.

The acquisition system

ANA is a terminological acquisition system based on statistical operations which automatically extracts terminologys from the text corpora.

This tool uses a statistical approach and uses a technic similar to the mutual information retrieval technic [Church & Hanks 1989]. It does not need any dictionary or any formal grammar. It works on no tagged corpora. Consequently, it can be used in a multilinguistic environment.

It analyses the most frequent word sequences in the texts and marks automatically the functional word (functionnal words belong to terms).

The process is incremental. Before processing the text, a bootstrap of a few representative words of the text is manually given. The first stage of the process acquires new terms which contain word of the bootstrap. For instance, if the word *tool* is included in the bootstrap, the term *machine tool* can be retrieved from the text, if it appears often enough in the text. Now, the term *machine tool* is included in the bootstrap for the next stage. The process continues incrementaly and stops when no more new terms can be retrieved. This system is able to retrieve terms written under different forms as, different inflected forms, uppercase or lowercase letters, and spelling mistakes. This means that a term is associated with a unique entity even if we meet it under different forms in texts. At the end of the process, the terms are validated by a human expert (a linguist).

The first experiment

The user queries the Internet database by using a search engine (for example LYCOS) [Lycos]. A list of references is then given to him by the search engine. The user can select on a subject of interest from the given references. ANA our terminological tool is then applied on the selected texts in order to extract their terminology, so the user, knows immediately if the contents are of interest or not.

Interest and limit of our demarch

The most important interest of our demarch is the speed of obtaining information about the semantical content of a text. Having information of the principal terms of a text gives us information on its contents [Jacquin & Liscouet 1996]. The user does not have to thoroughly read the text in order to decide if it interests him. These experiments have provided satisfactory results. The biggest problem encountered was the treatment delay in retrieving a datafile via Internet.

Our principal terminological extraction tool (ANA) uses statistical criteria. It gives good results if it processes a large scale of textual data. The HTML retrieved texts are a little bit small, we don't retrieve a lot of terms. We would have better results, if we worked with bigger text, for instance, texts which are put in an FTP (File Transfer Protocol) site.

To improve the retrieving process, it would be possible to use a semantical network of terms of the domain. In the first search process (with Internet tools), the user could proposed keywords which interests him. The use of a navigation tool in a semantical network would increase the keywords number. Then, we will retrieve more text which could be of interest.

The second experiment

Our purpose is to experiment a technic to improve the indexing process on the web. The idea is to use our terminological tool ANA in order to extract the terms of the text, (this extracted terms are considered as the principal keywords of the text). Our terminological tool ANA is statistical, and a term is extracted only if it appears enough in the corpus. So, we must build a large corpus of text from the same domain in order to use ANA.

To this end, we use an internet navigator tool CASIMIR based on the MOMspider tool [Fielding 1994] in order to navigate on the web and to build an homogeneus corpus of text from the same domain.

To this end, we use a specific navigation strategy. We begin the process with a specific text, we follow its hypertext links and we retrieve the linked documents. The process continues incrementaly (a search deepth is fixed)

We have experimented two strategys. The first which is a deepth search strategy and a second which is a width search strategy. In the two cases, we have built a corpus of 1 Mo. Then, we have applied our terminological tool on the two corpora, in order to extract their terms. The second experiment (with the corpus built with a width search strategy) has led to better results: the corpus is more homogeneus. But, we don't have extracted a lot of terms.

The problems met are that the built corpus is not enough homogeneus. In a web page, the anchor are pointer on other web pages. The content of these pointed web pages is linked to the first page content. When step by step, we retrieve new pages, these pages become less and less pertinent to the previous topic. So the corpus becomes less homogeneus. Our terminological tool ANA provides no so good results as expected because it is based on statistical methods and a term is only extracted if it appears enough in the corpus. Consequently, the corpus is not homogeneus enough

Conclusion

The first experiment is the more interesting and provides the better results. We could have better results, if we have worked with bigger texts, for instance, texts which are put in an FTP (File Transfer Protocol) site.

The second experiment seems not to be pertinent. It is impossible to build automatically an important homogeneous corpus (> 1 Mo) (on a same domain) by using simple technics as those we have experimented. The solution to build automatically an homogeneous corpus, seems to be the generalization of the use of the metatags included in the web page (written by the page conceptor), in order to inform the potential user of the semantical content of the page.

References

[Church & Hanks 1989] Church , K., & Hanks, W., *Word association norms, mutual information, and lexicography* . In proceeding, 27th annual meeting of the association for computationnal linguistics ACL'89, vancouver.

[Enguehard 1993] Enguehard, C., *Acquisition de terminologie à partir de gros corpus* . Colloque Informatique et langue Naturelle (ILN'93), Nantes décembre 1993.

[Fielding 1994] Fielding, T.Roy, *Maintaining distributed hypertext infostructures: welcome to MOMSPIDER's web* . in acts of the first international conference on the world-wide-web, WWW94, Geneva, 25-27 may 1994

[Jacquin & Liscouet 1996] Jacquin, C., Liscouet, M., *Terminology extraction from texts corpora: application to document keeping via internet*, proceedings of the TKE'96 international conference (Terminology and Knowledge Enginering), Vienne, le 21-23 août 1996.

[Lycos] LYCOS, Internet tool: www.lycos.com

Intranet IC Design Environment

Chau-Iou Jeng, National Semiconductor, USA
jeng@berlioz.nsc.com

Introduction

According to Dr. Andy Grove, CEO of Intel, Pentium will reach 10 Giga Hz with 1 trillion transistors by year 2011 comparing with today's 5.5 million transistors and 200 Mega Hz. The capacity of chip is ballooning as the deep sub-micron technology such as .18 micron is deepening. To achieve this goal, major semiconductor companies are exploring deep sub-micron technologies. The EDA (Electronic Design Automation) departments inside design houses try to drive Design Methodology (DM) with shareability and Intellectual Property (IP) portfolio reuse such that System-on-a-Chip (SOC) in 6-month Time-to-Market (TTM) can take place.

The Intranet IC Design Environment will deliver DM and IP block reuse in the real-time manner. It is been designed on a three-tier client-server model at National Semiconductor now. It may become an N-tier model in the future depending upon how server management is implemented and heterogeneous platforms are constructed. This environment entails three major components that are web front-end GUI (Graphical User Interface), server back-end management, and embedded networking communication middleware. The entire environment is implemented in java [Arnold et al. 1996]. For security concerns, it is designed on intranet base.

Three-Tier Client-Server Model

The three-tier client-server model is composed of web client, broker server, and execution server. The IC design could be fulfilled on any clients such as Network Computer, webTV, Personal Digital Assistant. It will result in the enormous budget saving on workstation upgrades. The broker server will cope with client requests and management on servers, and EDA tool licenses. Each request will be handled by a light-weight java thread [Lea 1997]. It can be constructed on top of LSF (Load Sharing Facility). The execution server can be a node inside an LSF cluster, or any powerful machine in the network without any LSF configuration. It will be appointed by the broker server based upon the job function and complexity.

Web Front-End GUI

Platform neutrality of IC design will come true with the confluence of internet ubiquity and java advent. A java-capable browser will be the only entry in the future IC design environment. The applet with front-end GUIs will open the door to enter the IC design world. It will present the graphical DM with EDA tool access. The DM provides the sequence of tool relations and data dependencies such that it can simply automate the tedious invocation for various EDA tools with lengthy library data. IC designers can simply follow the graphical design flow to invoke EDA tools with technology libraries, and view the results from various tools. Since the DM is probably design specific, configurations of various DMs are necessary. Hence, an embedded parser to deal with the configuration file will be a key engine in the web GUI applet. The DM is no longer a static document over the web. It can be dynamically verified and physically shared in tomorrow's environment. Also, the real-time graphical verification flow for IP blocks can be described in the different levels from behavior, RTL (Register Transfer Level) down to gate, even to GDS II file for fabrication.

Server Back-End Management

The server back-end software is emphasizing on management of servers and EDA tool licenses now. Server management will offer the best-fit machine to execute a job based upon the client's request. It consists of networking management, data access, and job loading analysis. In most cases, EDA license management would focus on the availability of license access. However, the mix of floating and node-lock licenses will be a major concern. In addition, the queuing mechanism will be another means to get late license access.

To execute the EDA tool with data access, the implementation is based on UNIX NFS (Network File System) in our environment. The integration with an open data repository of PCTE (Portable Common Tool Environment) [Wakeman et al. 1993] could be a major future enhancement. The entire management can be integrated with LSF (Load Sharing Facility). It will alleviate a lot of painful efforts.

Embedded Networking Communication Middleware

The embedded middleware sitting between the web front-end GUI and the broker server plays a key role in the development of client-server systems. It provides the following essential tasks [Buck-Emden et al. 1996] in this environment: isolate web GUIs from specific hardwares for platform-neutrality, provide open communication interfaces for distributed EDA tool invocations, control and monitor distributed transactions for EDA tool access, and furnish with object management functions for EDA tool classes.

For transaction control and monitor, the APIs (Application Programming Interfaces) allow the designer to monitor the status of tool execution in the web GUI, and control the tool running sequence in the server side. Basically, the web GUI contains a message window that can simultaneously display the status, when EDA tools are requested to run in the servers. The object management will stress on coupling minimization, cohesion maximization, and full inheritance and polymorphism between objects. It will mainly tackle the object relations on client sockets with port numbers, server sockets with threads, and EDA tools with licenses.

Conclusions and Future Works

This environment will provide the major IC or semiconductor corporations with a "platform-independent and customizable front-end" to current and future design tools. It will accommodate DM delivery, IP block reuse, effective EDA tools operation with minimal administration, and full networking resource utilization as well.

The future works will integrate with the internal work of Formal Verification [Beatty 1993], seek for partnerships with EDA vendors, pursue the collaborative workflow [Lavana et al. 1997] for concurrent engineering, and tackle the design process management [Sutton et al. 1996].

References

[Arnold et al. 1996] Arnold K., & Gosling J. (1996). The Java Programming Language, Addison Wesley Publishing Co.
[Lea 1997] Lea D. (1997). Concurrent Programming in Java: Design Principles and Patterns, Addison Wesley Publishing Co
[Wakeman et al. 1993] Wakeman L., & Jowett J. (1993). PCTE: The Standard for Open Repositories, The PIMB Associate.
[Buck-Emden et al. 1996] Buck-Emden R., & Galimow J. (1996). SAP R/3 System: A Client/Server Technology, A. W. Co.
[Beatty 1993] Beatty D. (1993). A Methodology for Formal Hardware Verification with Application to Microprocessors, Ph.D Thesis, Carnegie Mellon University.
[Lavana et al. 1997] Lavana H., Khetawat A., Brglez F., & Kozminski K. (1997). Executable Workflows: A Paradigm for Collaborative Design on the Internet, 34th Design Automation Conference, Anaheim, CA. 553-558.
[Sutton et al. 1996] Sutton P., & Director S. (1996). A Description Language for Design Process Management, 33rd Design Automation Conference, 1996. Las Vegas, NV. 175-180.

Acknowledgments

This work is funded by Nif/T committee at National Semiconductor. I especially would like to appreciate the Nif/T member Jewel Savadelis and my boss Dr. David Lin for their persistent support. Also many thanks go to Benjamin Hu and Ching Lai for their implementations, and Alvin Ling and Raheel Hasan for their valuable suggestions.

BIOME: A Scientific Data Archive Search-and-Order System Using Browser-Aware, Dynamic Pages

Sarah V. Jennings
Research Associate, Transportation Center
University of Tennessee, USA
xqj@ornl.gov

Teresa G. Yow, Ph.D.
Systems Analyst, Computational Physics and Engineering Division
Oak Ridge National Laboratory, USA
tgy@ornl.gov

Vickie W. Ng
Systems Engineer
Hughes Information Technology Systems, USA
vw5@ornl.gov

Introduction

The Oak Ridge National Laboratory's (ORNL) Distributed Active Archive Center (DAAC) is a data archive and distribution center for the National Air and Space Administration's (NASA) Earth Observing System Data and Information System (EOSDIS). Both the Earth Observing System (EOS) and EOSDIS are components of NASA's contribution to the U.S. Global Change Research Program through its Mission to Planet Earth Program. The ORNL DAAC provides access to data used in ecological and environmental research such as global change, global warming, and terrestrial ecology.

Web Pages Over Database

Because of its large and diverse data holdings, the challenge for the ORNL DAAC is to help users find data of interest from the hundreds of thousands of files available at the DAAC without overwhelming them. Therefore, the ORNL DAAC has developed the Biogeochemical Information Ordering Management Environment (BIOME), a customized search and order system for the World Wide Web (WWW). BIOME is a public system located at http://www-eosdis.ornl.gov/BIOME/biome.html.

Managing large amounts of data requires metadata, or data that describes the data, which is stored in a relational database management system. Several Sybase metadata databases form the heart of BIOME by managing to treat many different types of data in a consistent manner. Using metadata stored in a relational database management system allows for efficient searching of hundreds of thousands of records.

The data itself is stored on-line, off-line, and near-line. Small tabular datasets are stored on-line on spinning disk. CD-ROMs, tapes, and proprietary data are stored off-line. Larger datasets, i.e., satellite imagery, are stored near-line in a mass storage system. A browse capability allows users to preview near-line images by generating a thumbnail .GIF image of a larger imagery file. The location of the data is transparent in that the user does not need to know or care where the data is stored. With the exception of hard media (e.g., CD-ROMs) all data delivery is automated.

Browser Aware, On-the-Fly HTML Pages and Graphs

The ORNL DAAC WWW site categorizes browsers based on their capabilities. Pages are created according to the ability of the user's browser to display them. High-end browsers can get pages with frames, tables, and Java applets in addition to the information available to character-based browsers.

Because many of our users are scientific researchers working in remote areas, we must balance their needs with those of users who have access to the latest technology. On-the-fly HTML page customization allows the ORNL DAAC WWW site to take advantage of the most innovative WWW features while maintaining backwards compatibility with older browsers and text-based browsers. In a one year period, 1152 unique browser/platform combinations accessed the ORNL DAAC site. Because of the browser-aware on-the-fly page creation capabilities of BIOME, the DAAC was able to respond to this challenge by presenting each combination with customized HTML pages.

The ORNL DAAC's WWW site is designed around include statements that pull in the appropriate "modules" for each browser. If a user's browser is capable of displaying a certain feature the section of the page that uses that particular feature is included. If not, that portion of the page is not included for display. Thus, the page is dynamically altered.

BIOME allows users to see a graph of selected data. Tabular data are parsed according to arbitrary classifications describing the configuration of the data. The GD1 library is then used to generate a plot of the data. The user's browser is sent a .GIF with the selected labeled columns plotted in color. This technique allows one graphic engine to display many different layouts of tabular data.

WWW-based Tools

As the complexity of the DAAC's data holdings has increased, the task of maintaining the databases has become increasingly difficult and time-consuming. Fortunately, custom WWW-based tools make the task of the database administrator less difficult.

For example, the DBA Maintenance Tool handles the ingest of new metadata by providing on-the-fly templates of database tables generated dynamically from Sybase's system tables. New data can be typed onto the templates, eliminating the need for manually constructing Sybase bulk copy files, a task that is tedious and error prone. In addition, the DBA Maintenance Tool easily handles updates to existing metadata, offering such options as global updates to the databases. Other options include automated bulk copies out of the database and the printing of the current structure for each table. The DBA Maintenance Tool also automatically generates a transaction log that provides a record of all DBA actions on the databases.

Conclusion

The ORNL DAAC provides WWW access to a large number of ecological and environmental datasets. The DAAC has accomplished this task by designing and offering a customized WWW search and order system that allows efficient and rapid data search and retrieval. By developing customized WWW tools to manage global ecological and environmental data, the ORNL DAAC has made an important contribution to NASA's Mission to Planet Earth Program.

Acknowledgments

Research sponsored by NASA under Interagency Agreement DOE No. 2013-F044-A1 under Lockheed Martin Energy Research Corp., contract DE-AC05-96OR22464 with the U.S. Department of Energy. "The submitted manuscript has been authored by a contractor of the U.S. Government under contract No. DE-AC05-96OR22464. Accordingly, the

U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes."

Integrating Technology into Secondary Science and Mathematics Classrooms: A Partnership

Elizabeth T. Joseph, Ed.D., Assoc. Prof.
Secondary Education / Foundations of Education
Slippery Rock University
Slippery Rock, PA
elizabeth.joseph@sru.edu

Judy Pezzulo
Coordinator of Technology
Greenville Area School District
Greenville, PA
jpezzulo@greenville.miu4.k12.pa.us

With $5000 from a Corporation for Public Broadcasting / Ernest L. Boyer Next Step Grant, Slippery Rock University and Greenville Area High School (GHS) formalized a partnership. In order to apply for the grant, preparation included both faculty at the high school and at the university.

Preparation Informal assessment took place as discussions with various faculty at both institutions to determine where to begin with the task. In talking with GHS faculty, the areas identified as most ready to integrate technology into curricular practices mathematics and science. Once these areas were identified as the focus of this project, more in-depth conversations with these faculty members occured. Research was also done to familiarize the authors with earlier studies of computer use in these areas.

Identified needs Pre-assessment included informal discussions with faculty; this had identified four areas that the GASD math and science teachers felt were important focus:

* employing databases for student analysis of scientific data;
* using probing devices to explore scientific concepts;
* developing computer-controlled interactive videodisk programs; and,
* identifying and learning to use the best software programs to connect technology to curriculum.

SRU identified the need to place pre-service teachers (student teachers and field students) into technology-rich classrooms where they would have the support for technology grounded in school realities.

The Proposed Model

The proposed model developed from these identified needs consists of three basic identifiable components. Opportunities for face-to-face meetings would be provided. Communications avenues to discuss experiences with colleagues as well as to access information would be developed. There would be needed support for teachers to try at least one unit using technology.

Activities. Computer software and equipment were selected based on recommendations by the GASD faculty. Both Macintosh and Windows formats were purchased to meet the needs of both institutions. Software included: *Geometry Inventor*; *Green Globs*; *Mathematics Toolbox*; *Differential Calculus*; *Microsoft Works*; *Hyperstudio*; and *Internet Coach*. A variety of probes for the TI-92 graphing calculators were ordered. University personnel were recruited to meet the needs determined in the pre-assessment. Student teachers from SRU were recruited to student teach in math and science classrooms at Greenville Area High School. A "kick-off" luncheon was planned to give all participants from SRU and GASD the opportunity to meet each other and talk. During the luncheon, participants had the opportunity to air concerns and talk about what they expected to be included in the workshops. Six workshops were developed and implemented in February 1997. The workshops were:

1)Databases for Analysis;
2)Using the TI-92 calculator with probing devices to explore scientific concepts;
3) Computer Software for the Mathematics classroom;
4) Internet Sources for Math and Science;
5) Using Hyperstudio to develop computer controlled videodisk instruction;
6) Developing a HomePage.

All workshops were "hand-on." Participating teachers were expected to select one (1) area covered in the workshops and integrate it into their classes before the end of the school year. Although the workshops targeted the areas that were specified by the teachers, it was acknowledged that to "do learning" in a different way with new technologies, continued support would make the experience more successful. Once the workshops were completed, a listserv was set up to allow all participants to communicate with each other, sharing ideas, finding support for using information learned in the workshops in the classroom, and posting process and frustrations. This way, everyone participating at both sites could offer support, ideas, and encouragement. This would also provide the experience of learning to teach with the technology by using similar technologies. E-mail between the teachers and the specific "expert" was also encouraged. A listing of all e-mail addresses was given to each participant with the specific expertise of each listed. *Quick Cams* were purchased so that faculty at GASD and SRU personnel could continue face-to-face communications through the use of *CU SeeMe*. Times were scheduled for individual teachers to contact Slippery Rock faculty through the use of *CU SeeMe*. Every effort was made to provide easily accessible support for faculty involved in the project.

The last five weeks of SRU's semester field students from the university were placed in the classrooms of participating GASD faculty; here they observed technology being used to meet curricular needs. At the end of the semester, a final meeting of participants was held to discuss the results of the project and to look at the future of the partnership.

Results As a result of this project, there were several observable changes in the math and science classrooms:

- *Green Globs*, the equation graphing program was incorporated into math classes.
- Calculus class integrated the use of *Differential Calculus*, a CD-ROM, into some lessons.
- Multiple probes were used with the TI-92 calculators in the science classes.
- Also used in the science classes were Internet sites that were introduced to them in the workshop and others were discovered by the science students, notably *Scientists on Tap*.
- A HomePage was developed for the school district.
- A short *Hyperstudio* stack was developed for use with *Windows on Science*.
- Several student teachers from SRU had the opportunity to participate in and prepare a lesson plan using some form of technology.
- Field students from SRU had the opportunity to observe technology being integrated into the classroom.
- A learning circle was developed which involved a diversity of learning and teaching levels: pre-service teachers, in-service teachers, administrators, and university faculty.
- The use of technological means to continue the dialogue begun at the workshops also modeled using technology for learning.
- A formal, continuing collaboration between Slippery Rock University and the Greenville Area School District was developed.

The model used for the project provided initial learning for in-service and pre-service teachers as well as continued support to implement the integration of technology into the curriculum. However, at the end of the semester it became apparent that this was the authors' project not the project of all the participants. A sense of ownership by the participants was lacking which would have greatly increased the effectiveness of the project. Since this project is expected to continue, the model has been revised for the 1997 - 1998 school year.

Revised Model

The partnership between Slippery Rock University and the Greenville Area School District will continue to focus on providing a technologically rich classroom environment for both in-service and pre-service teachers. This will focus on the curriculum first rather than the technology. On-going support will continue to be available in this effort. Instructor level ownership will be developed early in the semester. This will be done by moving from the expert-learner design to a learning circle design by using engaged conversations / learning circles. Active partnership relationships will be sought; contractual or compensated time on a routine basis will be requested. Commitments will be defined continually.

Group reporting periods will be established with face-to-face meetings. The major purpose of these meetings will be for all participants to communicate with each other. Four meetings will be established during the semester with ongoing communications using the listserv, email, and CU SeeMe between meetings. Individuals will be asked to define individual goals before the first meeting. Considering those goals the first meeting will define semester goals for classroom activities while recognizing time constraints. Partner patterns will be explored and small learning circles will be established based on both sets of goals. In-service and pre-service teachers, SRU personnel, administrators from both institutions will participate in this meeting to develop partnerships.

This model will require more commitment from the participants and will demand continued communication. But this will also strengthen the integration of technology into the curriculum, continuing and improving the success of this partnership.

HANMAUM - A Multi-agent Model for Customer, Merchant and Directory Service

Dae-Ki Kang
Electronic Commerce Laboratory, Systems Engineering Research Institute, South Korea, dkkang@seri.re.kr

Joong-Bae Kim
Electronic Commerce Laboratory, Systems Engineering Research Institute, South Korea, jbkim@seri.re.kr

Ho-Sang Ham
Electronic Commerce Laboratory, Systems Engineering Research Institute, South Korea, hsham@seri.re.kr

Introduction

In recent years, some of classical researchers from the various areas such as Artificial Intelligence, Expert System, Network Computing, Parallel & Distributed Systems and even Computer Vision, have showed an interest in software agent and created several prototypes for them. Those software agents are to help users with e-mail and netnews filtering, Web browsing, meet scheduling, and online comparison shopping [Doorenbos et al. 1996]. In this paper we propose HANMAUM, a multi-agent model that performs scalable information retrieval, index scheduling and shopping cart control. In HANMAUM model, we have defined three entities, directory service, customers and merchants, and six different types of agents, a search broker agent, a meta-search agent, a demand-and-merge agent. an indexing assistant agent, a shopping cart agent, and a client managing agent. Those agents act autonomously to achieve scalable information retrieval, robot's visit scheduling, and reliable connection and state maintenance.

Background and Related Work

Scalable Information Retrieval

In directory service management, if the amount of the indexed information grows over more than that of secondary storage, the service manager should delete some of the index, or add additional storage devices. Also, existing directory services have no compatibility with one another for exchanging their information. Harvest [Bowman et al. 1994] is a good solution, but needs a fundamental change in index storage mechanism.

Robot's Network/Server Bottleneck

Search engines and directory services need index of other web sites for their service. The index of the web sites are made from the web pages taken from the sites. Robot automatically performs fetching, parsing, and breadth-first navigation. Robot's autonomous repetitive behavior causes serious server and network bottleneck.

Shopping Cart Problem

The fundamental reason for the shopping cart problem is that most HTTP(Hypertext Transfer Protocol) server mechanism is connectionless and stateless, which is not suitable for electronic commerce. A few HTTP clients can take care of states by cookies, for example.

Scalable Information Retrieval

The agents for the scalable information retrieval are search broker agent, meta-search agent and demand-and-merge agent. The search broker agent in HANMADANG communicates with other search broker agents on other homogeneous remote databases to propagate query and to find the result they have. A meta-search agent has a knowledge to request information to existing search engines, and he interacts with search broker agent. A demand-and-merge agent is at the customer's client to get the customer's query, select the proper search broker agent by light weight probe packet, send actual query and threshold, and merge the search result from the various search broker agent.

Indexing on Demand

An indexing assistant agent is installed on the merchant's system for indexing on demand(IOD). In classical indexing mechanism, a robot decides when to visit the web sites, notify his visit to the sites administrator in advance, and visit to download the web pages. IOD is to let the indexee(the sites administrator) decide when the robot should visit his site. He can also schedule the visit period of the robot to reflect regular update of his shopping mall.

Shopping Cart Preservation

In the view of electronic commerce, there is one critical problem to consider. Most HTTP servers and clients used in WWW now are stateless and connectionless, but, in online shopping, a customer's client and a merchant's web server should consider the way to know what the customer has done until now, whether the customer go out for lunch or coffee break without terminating his client, if the data displayed on the client is from cache, proxy server, or merchant web server, and whether the customer press stop button during the transaction or it is just a network failure. To solve these possible problems, it is "shopping session and history management" that is needed in online shopping. Shopping cart preservation is one of the common problems in electronic commerce. The fundamental reason for the problem is that HTTP(HyperText Transfer Protocol) server mechanism is connectionless and stateless, which is not suitable for electronic commerce. A few HTTP clients can take care of states by cookies, which is like a small size of shared memory between client and server. A shopping cart agent in the customer's side takes the role of preserving customer's shopping history and current states in the shopping mall. A client managing agent in the merchant's side is for control clients' shopping states and timeout.

Conclusions and Future Work

We have presented HANMAUM, a multi-agent model for directory service and connection management between merchant and customer with scalable information retrieval, index scheduling and shopping cart control features.

We have viewed actual electronic commerce in a simple way for now. We do not consider cyberbank, Certificate Authorities(CA), etc. in HANMAUM model. The further study is to include these entities to design a truly integrated model.

References

[Doorenbos et al. 1996] Doorenbos, R. B., Etzioni, O., and Weld, D.S. (1996). A Scalable Comparison-Shopping Agent for the World-Wide Web. *Proceedings of the first international joint conference on Autonomous Agents*, 924-929.

[Bowman et al. 1994] Bowman, C. Mic, Danzig, Peter B., Hardy, Darren R., Manber, Udi, Schwartz, Michael F., and Wessels, Duane P. (1994). The Harvest Information Discovery and Access System, *Proceedings of The Second International World Wide Web Conference*, 763-771.

INTEGRATING WEB-BASED SYSTEMS INTO TEACHING INSTRUCTIONAL TECHNOLOGY

Edward Kazlauskas
Instructional Technology, School of Education
University of Southern California
Los Angeles CA, 90089-0031 USA
E-mail: kazlausk@mizar.usc.edu

I. Soner Yildirim
Instructional Technology, School of Education
University of Southern California
Los Angeles CA, 90089-0031 USA
E-mail: iyildiri@ucs.usc.edu

Daniel T. Chan
HYNET Technologies, Belmont CA 94002 USA
E-mail: Daniel@hynet.com

INTRODUCTION

A portion of the master's program in Instructional Technology at the University of Southern California (USA) School of Education is offered through a cooperative agreement with the Los Angeles County Office of Education (entitled the Institute for Technologies and Learning - ITL) for the purpose of training specialists who will guide schools and related-type organizations in technology planning, implementation, utilization, and research.

Courses in the program are offered through the use of distance learning technologies such as e-mail, virtual libraries, multimedia institutes, Web-based chat rooms, and Web-based electronic books as well as through personal interactions with university faculty and Los Angeles County Office of Education professional staff. Significant parts of the program are designed for pursuit individually and in small work-place groups. Student assignments are organized to build on workplace goals and internship opportunities are available in technology oriented learning organizations. Successful completion is predicated on successful demonstration of six learner competencies: These learner competencies are:

1) Instructional design, use of technology tools and the integration of tools into teaching strategies.
2) Applying theory of human learning relative to the use of technology in instruction.
3) Designing and directing learning resources management.
4) Leadership skills in advocating roles of technology and information literacy in the reform of education.
5) Designing and maintaining infrastructure and connectivity.
6) Interpreting and conducting research in technology and learning.

WEB-BASED SYSTEMS

Various delivery mechanisms and Web-based systems are integral components of the ITL. These include the use of a Web-based digital library textbook and the integration of two Web-based projects into the curriculum.

The Web-based form of a textbook, Instructional Technology; A Systematic Approach to Education [Knirk & Kazlauskas, 1997], is organized into 14 chapters, a portion of which are used, for example, in the first course in instructional design. A student has the ability to navigate, read and review content, search, and annotate the electronic text. Other common Web features, such as copying and printing text, are also available.

The applications associated with two Web-based funded projects are integrated into the content of the instructional technology program, for example in the courses which deal with technology tools and technology integration. The project, Information System for Los Angeles (ISLA), is an exploratory regional information system for classroom integration of digital humanities materials funded by the National Endowment for the Humanities (NEH) and other organizations. ISLA provides access to digital research archives of Los Angeles materials in multiple information formats and its scope includes the widest variety of information from all historical periods, linked by spatial and temporal coordinates. The primary, long-term goal is to create a system that will enable all kinds of users, including K-12 students, to search and access a rich and diverse range research materials.

The other project incorporated into the curriculum is the Virtual Factory Teaching System (VFTS) funded by the National Science Foundation (NSF). This project addresses the educational needs of new engineers, and potential engineering students, by creating a manufacturing education workspace that will exist in the intersection of the three domains of education, the Internet, and virtual factories. The workspace takes advantage of advanced communication technologies in presenting manufacturing complexities in a realistic setting. The design of the workspace will enable students to participate in the functioning of the virtual factory by assuming the roles of various factory personnel in small team settings. Through acting out these roles, they will witness the range of decisions an engineer or a manager makes and their effect on the performance of a company. Student teams may even span institutional boundaries.

For both of these project, students are involved in learning the use of the applications, and then in integrating the applications into K-12 classroom settings through the development and evaluation of lesson plans and appropriate teaching materials.

CONCLUSION

One of the keys to the success of technology in the classroom is appropriate teacher training in technology use, integration, and classroom teaching approaches. The value of technology is limited unless technology training is integrated into the entire teacher education curriculum (Yildirim, 1997). To this end, we are incorporating technology both into the delivery of the instructional technology program, as well as integrating applications of technology, specifically web-based applications, into course content. The program is still new and we are currently in the process of examining the effectiveness and usability of the various Web-based systems used in the instructional program.

REFERENCES

[Knirk & Kazlauskas, 1997] Knirk, F. G. & Kazlauskas, E. J. Instructional Technology; A Systematic Approach to Education. Belmont CA: Hynet Technologies Inc. [Note that the textbook is login and password protected].

Yildirim, I.S. (1997). Effects of an Educational Computing Course on Pre-service and In-service Teachers' Attitudes Toward Computers. Unpublished Doctoral Dissertation. University of Southern California, Los Angeles.

ACKNOWLEDGEMENTS

The Web-based applications described in this paper are supported in part by Hynet Technologies Inc., and grants from the National Endowment for the Humanities (ED-20525-96) and the National Science Foundation (CLRT-CDA9616373).

The Web Lecture System (WLS)

Dr. Richard L. Klevans, rlklevan@eos.ncsu.edu
Department of Computer Science, North Carolina State University

Overview

The Web Lecture System (WLS) is a tool for constructing, editing, and managing Web-based presentations. These presentations consist of HTML documents with synchronized-streamed audio. A main component of WLS is an on-line editor that allows instructors to prepare slides for delivery. During live presentations, the system captures audio and timing data and automatically creates a Web-deliverable version of that presentation. To make the capturing and delivery process as simple as possible -- so that WLS can be used in a regular classroom setting -- all of the details of the underlying system are hidden from the users.

The WLS will allow students to view presentations on demand using a standard Web browser, such as Netscape™, and listen to the accompanying audio via a RealAudio™ player. Students are given several viewing options, such as fully automated paging, manual paging, or printable versions of the presentations. The system also has the ability to deliver live presentations with student interaction.

Description of Current System

An instructor accesses WLS through an on-line editing tool that manages class lockers. A class locker can be broken down into General Information, Configuration Information, Slide Sets, Lectures, and Access Rights. The *General Information* section contains information such as the name, phone number, and email address of the instructor. The *Configuration Information* section lets the instructor specify aspects of the Web pages that are generated by the system, such as which icons to display for the control buttons.

Slides Sets are used to specify groups of HTML files that are related together by topic. Due to the fact that instructors do not necessarily complete the discussion of a set of slides in one class period, a second data item, called a *Lecture*, is used to specify the slides (taken from the slides sets) that are to be used in any given class period's presentation. A class locker also contains a list of *Access Rights* to specify which users of the system are allowed to modify a class locker.

The WLS Delivery Process

Creating Slide Sets

When a new Slide Set is created, it is given a unique identifier and a directory on the HTTP server. The content, which includes HTML documents, images, and Java Applets, is usually placed in the Slide Set's directory; however, a Slide Set may contain links to content at other sites. The system will display the location of the Slide Set directory when the instructor edits a Slide Set so that the user knows where to place the content. Although WLS has facilities for creating HTML files, most instructors will find it easier to create HTML files with standard tools, such as Microsoft Word or PowerPoint, that have the ability to output HTML files. WLS has some features for making it easier to use the output of such tools.

The instructor must specify which HTML files in the Slide Set directory are part of the Slide Set and what the desired ordering of these files is. This list can be specified manually by typing the list of HTML filenames into the "Slides" editor. However, many of the HTML file creation tools output a series of consecutively numbered files, such as "slide1.html, "slide2.html", etc. If these files are placed in the Slide Set directory, then the "Make Slide Names" command can be used for adding all of the filenames to the Slide Set, instead of manually typing in the names into the "Slides" editor. Finally, WLS can convert an HTML file with horizontal bars as page breaks (i.e., the <HR> tag) into separate HTML files - one file per page.

Creating Lectures

When a new Lecture is created, it is also given a unique identifier and a directory on the HTTP server. A Lecture contains a list of slides taken from the Slide Sets that have already been created. Individual slides are specified with the following syntax: ss*m*:*n*, where *m* is the Slide Set identifier number and *n* is the slide number within that Slide Set.

The list of slides for a Lecture can be specified manually by typing ss*m*:*n* entries into the "Slide" editor or systematically created by using the "Use All Slides From" and "Use Subset From" commands. These commands will automatically create a list of ss*m*:*n* entries based on a range and a Slide Set identifier. Additionally, entries can be created by using the *Slide Chooser* which allows the instructor to visually inspect slides before adding them to the lecture.

Presenting Lectures

When a live presentation is given, a Web version of the lecture is used as the slides for the presentation. The slides can be displayed to the audience via an LCD projector. Naturally, the computer used for displaying the HTML pages must either be connected to the network or be a standalone HTTP server.

When the instructor starts a presentation, the RealAudio Encoder, the program that captures audio information, is automatically started. The instructor will use "Next", "Previous", and "Done" buttons to navigate through the presentation. As this navigation occurs, the HTTP server will record the times in which each slide is viewed in a data file for the presentation. When the presentation is complete, the encoder is terminated and the on-demand Web-deliverable version of the lecture is automatically created.

Additionally, students can connect to a lecture in progress and receive the streamed audio and synchronized HTML files, as the instructor is delivering them. This allows remote students to interact with the instructor during the lecture. Feedback from the remote students is delivered as text messages with immediate notification given to the instructor when new messages arrive. Furthermore, an electronic whiteboard can be used for capturing on-the-fly drawings or annotations of images.

Future Goals

As network bandwidth increases in the future, we will be able to incorporate streamed video into our system in order to capture additional information from the original presentation such as gestures and facial expressions. One possible solution is to use MPEG-2 compressed video over Asynchronous Transfer Mode (ATM) networks. We are also in the process of incorporating the RealVideo^{TM} technology into WLS.

Conclusions

In summary, we have discussed a system for easily converting live classroom presentations into low bandwidth, Web-based versions of the same presentations that are available on demand to a wide audience. WLS is currently being used in university classroom and industrial training settings. More information on WLS is located at: http://renoir.csc.ncsu.edu/WLS/

Acknowledgments

This project is supported, in part, by the NSF awards ACS-9418960, ACS-9418606, and ACS-9696131. RealAudio™ and RealVideo™ are trademarks of Progressive Networks Corporation. Netscape Navigator™ is a trademark of Netscape Communications Corporation.

RISK-1: its way to the Virtual Reality (Taking risk in virtual environment)

Vladimir Kosolapov, Ph.D.
Head of the Department
Institute of Mathematical Machines and System Problem
National Academy of Sciences of Ukraine
tel: 38-044-2660786
fax: 38-044-2664091
Email: terem@immsp.freenet.kiev.ua

The goals, objectives and policy measures required by the European Community entering the 21st century must be examined in the context of Globalisation. Nowadays social, political and economic processes are especially urgent in our society. All members of the community watch with keen interest the economic and political events, the course of reforms and transformation processes which take place on the global level. They take part in referendums, election campaigns, meetings, strikes and so on and so forth. Social and economic restructuring is connected with decisions made on cultural, national, political,economical, technological, environmental issues.

Policy makers, economists that carry out reforms, state figures who make political and economic decisions, legislative and executive bodies are permanently in need of up-to-date information-analytical technologies and techniques of comparative analysis, methodologies of building and studying the dynamics of interests of main actors group both for internal and international relations with the account of global issues [Beltrami 1977; King 1990]. Given technologies of the comparative analysis ensure the possibility of an efficient preparation of urgent political and economic decisions and draw highly-qualified experts for working out, analyzing the most pressing problems of the political life in the society and the public opinion appraisal. A virtual decision-making model for a social system model can provide a powerful tool for Internet users, especially for those in need for the analytical management support. A variety of applications is due to the proposed hierarchy of different participants including ethnic, social, economic, political and other groups. Correlation of forces, their influence on the political issues achieving the balance of interests is used to estimate the undesired consequences of different activities.

The model can be equally applied for both the geopolitical and economic monitoring. Its basic stages include: preparatory,information, analytical, multifunctional modelling and summarizing of the materials received.

Advanced data bases and knowledge bases, the decision-making support systems, expert systems serve as a foundation for social-political and economic technologies, ensure information survey, its reliability and urgency, the possibility to compare opposite points of view for solving political, social, economic and other problems. They provide ways for searching trade-offs and how to gain an adjustment of these problems. Alongside there is a possibility to use flexible means of communication: computer networks, telecommunication, advanced computing and informational techniques on the global scale.

The growth of up-to-date analytical techniques, methodologies, informational technologies allow to set up and realize the following components of political, economic reforms and transformation as systemic approach, possibility to forecast undesired consequences of decision-making and to assess the level of social tensity, to reveal zones and factors of risk in the corresponding geopolitical regions, strategic areas, spheres of common interests [Wolf 1988]. Therefore a systemic geopolitical and economic monitoring becomes a pressing problem.

In our project we try to make an appraisal of:

- the public, economic and political situation (in a country and between countries) and to predict changes set at a definite period of time (with indication of possible political and economic events on the global scale);
- the advisability of practical steps during the period of the economic transition (to carry out economic reforms, to analyze development programs, to form the budget, to make investments, to grant credits, to organize joint ventures etc.). Here you should take into account possible changes that can occur in a certain political situation and that might influence noticeably the efficiency and quality of this activity (the possibility to anticipate positive results etc.);
- abilities to control the situation by a political leader or a person responsible for economic reforms;
- the advisability to make crucial decisions on social and economic programs with regard of the political events that are likely to happen and are consequences of such decisions in the process of globalization.

The approach for building and studying the interests dynamics is based upon information-analytical systems that are able to assess events development and to prevent undesired changes of the political and economic situation during a process of transformation. Here it is supposed to make use of the techniques for measuring, assessing and giving preference to the alternatives that have mutual interest.

The degree and nature of the appraisal of the actors participation in all events will allow to express the major social stereotypes and priorities. The reconstruction of the transformation chain of social stereotypes and behavioural priorities which have been taken place during recent years will be performed with the help of the behavioural simulation while studying focus groups. Analysis of the dynamics and interconnections of the behaviour priorities within the social situation in a country will allow to receive data concerning the cause and effect connection between the development of social stereotypes, behaviour priorities, economic and political reforms in Ukraine (see Wolf, 1988).

Qualitative appraisal of the actors' "force level", calculation of the interests' line will be realized by using an up-to-date informational technology of a political analysis, combined with the expert system RISK-1 [Tikhomirov, 1981; Kosolapov, Morozov, 1993]. This system provides an original social system model. This model involves the hierarchy of the main participants (religious,ethnic, economic, political groups etc.) and correlation of their forces (estimation of their effect is realized with the help of the original techniques and algorithms). At the same time there is a possibility to achieve a balance of interests for the decision making. RISK-1 System user is provided with an updated information technique for political and economic analysis.

We are engaged in the development of a social system model for the analytical maintenance management. This model involves the hierarchy of the main participants of a social system (religious, ethnic, social, economic, political groups etc.), correlation of their forces and their influence. At the same time there is a possibility to build an interests' balance (in economic, political and social sphere) for the decision making and to estimate undesired consequences of activities, especially the chances of political and economic risk. The diagram below (see Figure 1) reflects, in general, an example of the primary stage of the decisions structure preparation that applies the above-mentioned model.

The conceptual model of the information and analytical support of the decision-making while carrying out the systemic geopolitical and economic monitoring suggests the following five steps: preparatory, informational, analytical, multifunctional modelling and summarizing of the materials received.

Preparatory stage that comprises the task generation takes into account customers' interests and aims; here one can determine the main actors of the events and their hierarchy; to define main packages of problems and alternatives while making decisions in political and economical spheres; to establish conceptions and requirements for the informational structure to form and load the database, to make sociological and experts survey, the economical analysis etc. It is supposed that for the confirmation of the initial hypothesis the attention will be also paid to the life standards indicators, the role of social stereotypes in the dominant activity of actors.

An informational stage includes the data collection and processing with the help of informational technologies; one can check the validity of the acquired data, to summarize experts' rating; to modernize databases with the purpose of forming multifunctional information environment, to make authorized programs and systems.

An analytical stage comprises the selection of the actual information; one can evaluate events, forces and the level of influence of the events' participants; to reveal urgent problems and conflicts; to make political and economic forecasting (to define the most probable ways of the situation development) and to appraise the level of the manifestion of undesired events in political and economic spheres. The corresponding diagram of the actors' degrees of influence is shown in Figure 2.

While doing an efficient comparative analysis of the situations the main actors are examined at different levels of the hierarchy of social divisions(depending upon the aims of the research or set tasks): religious and ethnic, public and political, social and economic, age and gender and so on. It is rather interesting to distinguish the main factors on the basis of which the degree of influence of the main actors at the geopolitical map of the region, country can be noticed. These factors can be the following: the quantity, the standart of life, the political activity, the creative potential, the represenation in the power bodies, the moral, legal and culture level, the contribution to the national income, the role in the economics management and a number of others [Davis R., Smith B. 1989; Feinberg 1985]. Each of the social divisions is associated with specific indicators which are connected with the conditions of the concrete situation. One of the possible versions is shown in Table 1, where the corresponding relative weights reflect each indicator.

The multifunctional modelling stage allows to estimate the chances of the purposeful changes of the level of risk in the decision-making - provides the risk management. The comparative analysis of influence factors based upon the information-analytical technologies allow to give a qualitative appraisal of the degree of influence of each actor upon the situation. At this stage positions of the main actors are defined. These positions may be represented as a numerous set of their desires, requirements, actions. The tolerance of their behaviour while solving exposed packages of problems, the analysis of life priorities are also of great importance. Here it is possible to examine the attitude of actors to alternative solutions of actual problems and to reflect objective and subjective contradictions and their consolidation.

The following problems are distinguished:

- the contradiction of interests when the parity of forces exist;
- the contradiction of interests and available possibilities;
- the contradiction of the priority character when several urgent and significant problems exist.

Each problem reflects a certain way of distribution of a limited number of material and spiritual resources and is expressed by the degree of satisfaction (or unsatisfaction) of the actors needs. Therefore while analyzing an actual situation expert groups first of all focus their attention on the actors interests and degree of their satisfaction for exposing urgent or potentially pressing problems or paticipants actions.

While analyzing the actors attitude to problems that comprise blocks of packages we build interests lines and compare them. Data on power distribution is also useful during the comparative analysis of the actors impact and their inclusions/exclusions in the society transformation because it helps to examine the lines of interests more effectively (see Figure 3). Here are some lines of interests which reflect the desirability of various alternatives from the positions of different actors. The most urgent point is to make an appraisal of the level of tensity and definition of security zones while having contacts and elaborating decisions on mutual problems. In case the contradiction exists between the actors then the balance of interest is defined.

At the last stage documents for decisions support are compiled. Different prospects of the situation, its tensity and changes are estimated. This stage of systemic geopolitical and economic monitoring gives an expert-analytic the possibility to build charts on the basis of an available information as regards the main actors, correlation of forces and orientation of separate problems and their alternative solution.

These charts are used by experts who analyze political and economic risk - "decisions structure". Such charts show the structure of the public and political situation, its economic aspects, expose main moving and conservative forces, define actors' interests and positions (see Figure 4).

While doing geopolitical and economic monitoring the comparative analysis involves standard computing procedures that allow to obtain probability appraisal of the risk degree while realizing the main hypotheses and scenarios of political decision-making [Quade 1989]. Alongside the user gets at his disposal the information about globalization processes in the geopolitical region that is under analysis, about problems that reflect these processes, about main actors and their influence upon the situation and also possible ways and scenarios.

These models are implemented in the DMSS "RISK-1". The system user is provided with an updated information interface of political and economic analysis. It makes possible to carry out commercial and political activities analysis under the more favourable conditions. For instance, to make profitable investments, stabilize political situation. The system was used for the forecasting the results of the economic and social developments in the Ukrainian regions; it was also applied for the information support and the forecasting results of the referendum on the issues of the Ukraine's independence and the first Presidential election in 1991.

The Virtual Reality for Sociological research is designed conceptually. The main components, structure, main actors, their interests are presented. At the present moment we are searching a suitable programming language for the reflection of this conception in the proper machinery. In case someone has suggestions - we ask to send them to our address.
We hope for further fruitful cooperation.

The further development of the RISK-1 project suggests transformation into Virual Reality the antholoqical approach to the construction modelling, as well as images and algorithmic calculation applied in this system.

Literature:
Beltrami, E.J. (1977) Models for Public Systems Analysis. London: Academic Press.
Davis R., Smith B. (1989) Expert Systems: how far can they go? // AI Magazine. - 1989. - 10. No 2. - P. 65-76.
Feinberg, S.E. (1985) Large-scale social experimentaion in the United States, in A Selebration of statistics (A.C.Atkinson and S.E.Feinberg, eds.). New York: Springer-Verlag.
King, P. (1990) International Economics and International Economic: A Reader: San Francisco State University.
Kosolapov, V., Morozov, A. (1993) The Basis of Geopolitical and Economical Monitoring. // Control Systems and Machines. Kiev. - 1993. - 4, P. 20-25.
Quade, E.S. (1989) Analysis for Public Decisions. North-Holland, New York, Amsterdam, London. 409 p.
Tikhomirov V.B. New International Development Strategy: A System Analysis Approach. - New York: UNITAR, 1981. - 32 p.
Wolf, C.J. (1988) Markets or Governments: Choosing Between Imperfect Alternatives. Cambridge, Mass.: MIT Press.

The Initial Reaction of Users to CALLware

Eva Fung-kuen Lai, Independent Learning Centre,
The Chinese University of Hong Kong, email: fungkuenlaw@cuhk.edu.hk

Introduction

In a city with no natural resources other than the 6.6 million people, in a city supported by service industry, in a city known to the world as an international financial centre, development in inter-personal communication skills is crucially important. When there are 150 pages of job advertisements on a Saturday in a local English newspaper, the job interview process has received much attention from final year undergraduates. On examining the market, however, there seems to be very few tailor-made learning materials either in the form of CD-ROM disks or video tapes to help interviewees to perform at their best in job interviews. Thus a CALL package is produced based on analysis of real job interviews in the Hong Kong setting. The package is used in the Independent Learning Centre, a self-access language centre of the university. In this paper, the computer-human interface issue will be highlighted for discussion, especially the way feedback from learners is to be used to improve the CALLware.

The CALL Package

Having collected data from students right after their real job interviews, eight episodes were written using the same job advertisements and job application letters. Teacher designed exercises, explanation, quizzes and comments linked to various parts of the interview dialogues were also written. There are11 icons on the first page, Preface to dialogues, Dialogues 1-8, Feedback and Project team members. Learners can go to any of the dialogues as they wish and they are invited to give feedback to the project team who can then improve the CALLware. If they press the Feedback icon, they will see blank areas for them to write under these headings: Content, Usefulness and Others. Feedback written will automatically go to the team leader's email address.

The Feedback

So far, more than 100 users have sent feedback to the project team and most of them are positive--describing the program as interesting, informative, useful, practical, well-organized and attractive. An example of such feedback is as follows:
"This homepage is very useful for those who are preparing for a job interview. It provides eight pieces of sample dialogues between interviewers and interviewees, each piece focusing on a different situation. Users are allowed to read each exchange and click on them to listen. Questions are provided afterwards to draw the users' attention to each point of significance that should not be ignored. Suggestions are given afterwards commenting on users' answers.
Each dialogue has its emphasis. Dialogue One is an introductory piece depicting the logic of conversation, telling us what interviewers expect and interviewees usually do. In Dialogue Two, situations are more complicated, there are more than one interviewer. Dialogue Three shows us how to deal with interviewer's follow-up question. Dialogue Four, Five and Six teach us how to handle unfavourable conditions such as difficult questions and unfriendly interviewer. Dialogue Seven serves as an example of a successful interview, while Dialogue Eight an unsuccessful one."

Other favourable responses are:
"I think the front page is very cute, the dolls are simple, lovely and colours are sharp as well."
'Those cases give us a real interview feeling, then there are the multiple choice questions and comments. It seems that we are really attending the interview and we can have our own answers."

"The dialogues include not only good examples but also bad ones so that we may pay attention to those inappropriate manners and avoid them."
"In the quiz part, it corrected me some of the misunderstanding about the job interview."
"The summary part gives us more information on matters concerning interviews through filling in the blanks and it is interesting."

Apart from encouraging remarks, users also told us what they wanted. The following extracts show what they would like to add to the package:
"group discussion (employer to a few interviewees) because I've heard that this sort of situation do exist in the real world and I really have no clue in dealing with this situation. I don't know whether I should show off and express all my abilities or just to be polite to give a better impression to the employer"
"Since I am a Science student, I would find some difficulties during interviews when the interviewers ask me whether my knowledge is suitable for the job which is completely different from my studies. Three of the dialogues involved students majoring in Business Administration. I would prefer to have some dialogues of interviewees who are from different faculties."
"In my experiences of interviews, I found that all interviewers asked me to introduce myself, and I am not sure what I should say -- my study, my family or other aspects. So I would like the dialogue to include this part."
"Include some concrete tips on interview skills such as some DOs and DON'Ts."
" Add a section that allows participants to send in their questions for corrections or comments."
"More explanation of vocabulary would be helpful."
"More examples on different occasions in which we may come across in an interview."

There were also negative feedback, for instance:
"The movable gifs at the top left corner very annoying and disturbing."
"Words are clustered together and it's difficult to read. Use double-line spacing."
"The front page doesn't indicate what the user should do next."

Discussion

In general, users like the content and the presentation of the CALLware. As they pointed out repeatedly in their feedback, and as revealed by an earlier market survey, the present CALL program has very neatly fitted in the gap in job interview training skills in Hong Kong. Users find it helpful to learn more about job interview situations, what is expected of them, how they should handle tough questions, what they should say if they can't answer the question, etc. The fact that the episodes were written based on real situations makes the CALL program more relevant to their needs. But as a computer program can't create and generate answers to questions as human beings can, it is even more important to rely on user feedback to improve the program to anticipate questions from users. Feedback collected has the started the second phase of material development adding more dialogues to help Science students.

Teachers at our centre also provide job interview practice to final year students before they go for the real one. But very often, they find that students make the same mistakes and they find it frustrating answering the same questions again and again. Some students try to get better prepared by watching videos or reading books based on business settings in America or Europe and they do not know how to react to situations when interviewers are all Asians or when the Hong Kong scene is the focus. The CALL program fits the learners' needs very well only if they knew of its existence.

After the program was produced, it took a long time and a lot of effort to promote it among students. Final year students still favour the workshops and the face-to-face practice and they still want to ask the questions that bother them even though the teachers have to answer the questions repeatedly. We ask teachers to introduce the program in their course work, we ask colleagues in the Appointment Office to tell final year students about the program, we send out posters and flyers and we talk about the program in our job interview workshops. For students not familiar with using the computer to learn, we offer CALL workshops to assist them in getting to the website.

Conclusion

Users like the CALLware as a whole as the content suits their needs. When the program cannot satisfy their needs, they tell us so and we continue to develop more. In this way, the CALLware is growing in response to learner needs and it has a life of its own.

Interactive Shared Bookmark

Luigi Lancieri
France Telecom
CNET de Caen, France
E-mail luigi.lancieri@cnet.francetelecom.fr

Introduction

ISB has been designed to looking up information on Internet easier. This tool aims to be operated in conjunction with local area networks, for instance inside a company or, more generally, within Intranets. ISB's capabilities are situated between bookmark's ones and search engine's ones. So to speak, ISB allows each user to take advantages of their colleague's bookmarks. Inter activity is one of the main functionality of this shared bookmark.

As in most companies, people are connected to Internet through a Proxy-server, for reasons regarding security and cache-provided benefits. ISB will make up a database from HTML pages that are present in the cache of the proxy-server, and this database is what is first consulted by users.

Company members have to use a traditional web browser to connect to ISB (on local HTTP server). An HTML form is available to them to formulate their queries using keywords similarly to traditional search engines (e.g. Alta vista, Lycos). ISB searches its database and then sends back to the user the addresses of distant sites containing information related to a specific topic, and ranked according to its level of interest.

Information pages can be viewed (by clicking on their address) and rated (good, average, bad) by the user. Such a rating makes it possible to sort addresses according to their level of interest. ISB also elaborates statistics on HTML page access rate, such statistics being taken into account in order to classify addresses.

ISB is suitable to multi-site companies, especially in Intranet context. To do this, ISB contains slave modules, which are scattered among the various company's sites, as well as a main module that manages information sharing between slave modules. Such an architecture makes possible a large information sharing between users of various sites. Nowadays, tendency is, indeed, to distribute caches in a global network architecture (e.g Squid project). The purpose is mainly to reduce web traffic and speed up the load of HTML pages. In this context ISB seems to be an efficient way to share between users the large amount of data contained in these different caches.

Architecture

The components of such an architecture are: a Proxy-server, an HTTP-server, a specifically designed software. This software can be reached by means of the HTTP-server through the Common Gateway Interface (CGI).

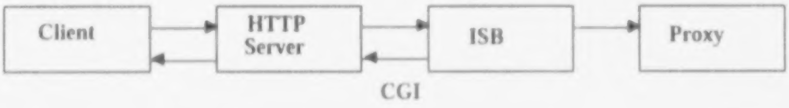


Figure 1. General architecture

ISB components implement the following main functions : the cache and proxy's log files' explorer, the dialogue with user and the building and management of the database.

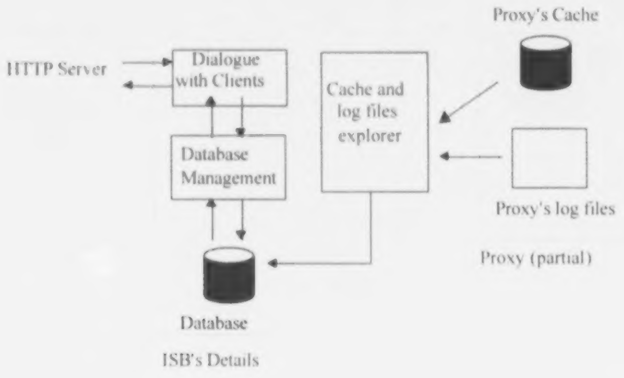


Figure 2. Detailed Architecture

Advantages

As said before, ISB should be considered as a groupware tool half-way between a local bookmark and a classical search engine. Using this technique presents several advantages:

1) Users receive results ranked according to several suitable criteria : an information quality rating granted by users, an access rate to HTML pages and the content rate of keywords per page. The subjective characteristics of some of these criteria, as information quality grade granted by users are corrected by averaging each criterion. Moreover the compound use of all of these criteria allows a good estimation of the interest of a certain page.

2) Everyone in a site can take advantage in their colleagues' searches, as all viewed pages are present in proxy's cache and are used to build the search database. This database is built according to requirements of a limited user's group (firm's staff), which makes the search process more accurate, so that the database size and the updating delay can be minimised. ISB is also expected to transfer unsatisfied requests to a « general » external search engine (e.g. altavista).

4) ISB is locally implemented, so that it can be reached easier and faster than a classical external search engine would be. The proxy server is set up to make a nightly updating of its cache. The desired HTML page is downloaded (if it is present) from the in-site proxy's cache, so that users have little time to wait.

5) Database building may be cheap, as ISB needs not run all over the web, unlike a classical in site or external search engine.

A large part of this system's interest and efficiency lies in its inter-activity. Actually, the more ISB is used, the more its database enriches itself (by learning process), and the more powerful ISB becomes.

Acknowledgements

BEST COPY AVAILABLE

All my thanks to Pierre Faurant for his involvement in the project developments.

Distributed Administration–Service Integration on the World Wide Web

Carlos Laufer
Computer Science Dept.
Catholic University of Rio de Janeiro
Brazil
laufer@inf.puc-rio.br

Hugo Fuks
Computer Science Dept.
Catholic University of Rio de Janeiro
Brazil
hugo@inf.puc-rio.br

Introduction

This research is oriented towards the development of a conceptual model of service integration using different types of media. Our emphasis is on the cooperative aspects of the group work. In this paper we present a prototype for distributed cooperative administration of a public CU-SeeMe reflector. Videoconferencing has been around for quite a while, but has only been available to large corporations that could afford tens of thousands of dollars in expensive equipment and proprietary high bandwith networks. Fortunately the situation is changing with the introduction of CU-SeeMe. The CU-SeeMe reflector is a software that works as an audio and video server, where all the clients connected to it receive everyone else's video and audio according to their preferences.

The Rio Internet TV Reflector

The Rio Internet TV reflector (RITV) at the Catholic University of Rio de Janeiro is the oldest public reflector in Brazil. It started its operations in early 94. It is a mature reflector receiving almost a hundred visitors a day. It has been cited in popular magazines and newspapers ([Ediouro 1996], [Mandarim 1996] & [Informática 1996]) and has a companion web site (http://www.inf.puc-rio.br/~refletor) which is a major source of information for the Brazilian CU-SeeMe community.

The RITV is listed worldwide and, as mentioned above, it receives a considerable amount of visitors daily. A new medium fosters unprecedent behaviour and a research (G-rated) reflector needs to be looked after most of the time. For that purpose, RITV administrators invited some of its regular and trustworthy users to be part of the reflector's daily administration.

In order to participate in the reflector's administration process, all the members of the administrative team should know about things like Telnet, passwords, reflector port command language and Unix knowledge, that are the ingredients related to the services needed to manage the reflector. These restrictions together with the reflector's twenty-four hour availability turned the reflector's administration into a difficult one–only highly specialized people could help in the job.

Ideally an interface should be offered to the administrators, that could hide the characteristics of each of the services previously mentioned from them. For example, in order to know who is currently connected to the reflector, one could just click a button instead of telneting the reflector's command port and typing the command *who*.

Another important thing that we considered was the ability to administrate the reflector without the need for any specific software, platform or location. This matches exactly with the client-server WWW model [Rice at al. 1996].

The Prototype

Our prototype comprises one html page and a CGI program. This program is denominated CGI Services Manager because it integrates all the different services needed to manage the reflector. To start the administrative process, first an entrance page–the only not on-the-fly page–is accessed by the administrators. The site is password protected, and different passwords will position the user in different levels of administration. After entering the password it is submitted to the CGI Services Manager which checks its validity and generates the main administration page, which lists who is currently connected to the reflector.

Clicking on the RITV icon, a new web browser will appear showing the companion site mentioned above. The *mail* item is used when an administrator wants to email another administrator or the collective of administrators. The *urgency* item brings a page that is the WWW interface to a pager system. It is very useful because the pager belongs to the person that actually does the *dirty programming job*.

The execute button refers to the five actions listed in the area to its right. The DNS option execute a lookup on the IP address of a selected participant. The other options are to kill somebody (takes the user out of the reflector), to deny (a persistent kill), to terminate the reflector–which is followed by a page designed to restart the reflector with a few configuration options. Finally the log option which presents the log of activities (kill, deny and reflector termination) generated by the administrators.

A fraction of the users that are taken out of the reflector write to the administrators here at the university, complaining about administrators attitudes. In these situations the log page is proving itself as a valuable source for conflict resolution. Together with the appropriate part of the chat–CU-SeeMe has a Chat window where reflector participants exchange lines of text–that took place between the offended part and the administrator, it is possible to reconstruct the conversation and to either confirm the administrator's action or to condemn it for his prepotence. To simplify the administrator's tasks we are coding a set of rules where the intention is to help the administrator when he approaches the user. The behavior expected from the users is what is called *netiquete* in a G-Rated digital community.

Conclusion and Future Work

This paper briefly reports the distributed cooperative administration of a CU-SeeMe reflector. Being a twenty-four hour public reflector, different people, with different Internet habits and within different time zones, were invited to help supervise the reflector. Differing from White Pine which has just released a reflector manager that works on a PC running NT operating system, we chose to use WWW pages to be the control panel for its administration because of its platform independence, ease of maintenance and facility of service integration.

At the moment the prototype is not making use of a database to store logging information. We plan to use one in order to query the database for helping the administration and for statistics.

Our main objective is to carry on the this prototype's development, trying to embed mechanisms into it that will further cooperation between our administrators like becoming aware if there is another administrator on board at that moment [Palfreyman and Rodden 1996].

References

[Dorcey 1995] Dorcey, T. (1995). CU-SeeMe Desktop Videoconferencing Software: CONNEXIONS, 42-45.

[Ediouro 1996] Ediouro (1996). Vejo você, você me vê - Um tutorial completo do CU-SeeMe: Guia da Internet.br, Ano I, N° 3, Ediouro S.A (Portuguese).

[Informática 1996] Informática *etc* (1996). Dá uma passada no meu monitor qualquer hora: O GLOBO, September, 23 (Portuguese)

[Mandarim 1996] Mandarim (1996). Janelas Indiscretas: BBS Magazine Ano 2 N° 13 Editora Mandarim (Portuguese).

[Palfreyman and Rodden 1996] Palfreyman, K., & Rodden, T. (1996). A Protocol for User Awareness on The World Wide Web: Proceedings of ACM 1996 Conference on Computer Supported Cooperative Work, Boston, 130-139.

[Rettinger 1995] Rettinger, L (1995). Desktop Videoconferencing: Technology And Use For Remote Seminar Delivery: M. Sc. Thesis in Computer Engineering, North Carolina State University, Raleigh, North Carolina.

[Rice et al. 1996] Rice, J., Farquhar, A., Piernot, P., & Gruber, T. (1996). Using the Web Instead of a Window System: Proceedings of Computer-Human Interface, 103-110.

Acknowledgments

This work was partially supported by CNPq: Carlos Laufer grant nº 143296/96-5 and Hugo Fuks grant nº 524557/96-9. We thank Professor Carlos José Pereira de Lucena, Executive Secretary of the Management Comittee of the Brazilian Internet–who is making the Internet a valuable asset for the Brazilian computing community.

Teaching College Mathematics with the WEB, Scientific Workplace, and Scientific Notebook

Lawrence E. Levine
Department of Mathematical Sciences
Stevens Institute of Technology
Hoboken, NJ 07030
llevine@stevens-tech.edu

Introduction

The second year mathematics courses at Stevens Institute of Technology deal with ordinary and partial differential equations, linear algebra, multiple integration, and surface integrals. During the 1996-97 academic year the author used Scientific Workplace (SWP) and Scientific Notebook (SNB) as tools in teaching these courses to 260 students. SWP and SNB are technical word processors that produce .tex files. They each contain a Maple kernel that allows the performance of a large number of mathematical procedures such as algebraic manipulation and simplification, graphing, differentiation and integration, solving algebraic and differential equations both exactly and numerically, matrix manipulation, etc. The World Wide Web was used as a vehicle for transmitting software and files as well as a learning tool. Several projects were prepared in SWP and SNB that required the student to use Maple. One project dealing with a mass-spring-damping system is *interactive* and represents a rather striking balance between analytic solution, the use of Maple, and simulation. A second project combines Web searching, software downloading and installation, and SWP or SNB to study some first order differential equations. Others deal with matrices, Fourier series, numerical solutions for first order differential equations, and multiple integration. This paper discusses the benefits gained by integrating computer technology into these courses as well as the problems encountered.

The Course in Differential Equations

Differential equations is a first semester sophomore course at Stevens Institute of Technology. It covers the standard topics dealing with first and second order ordinary differential equations. The course meets four hours per week with two hours of lecture and two hours of recitation (drill). The author lectured twice a week to 150 students divided into two groups. Scientific Workplace was the software used in conjunction with the text. SWP was distributed to students living on campus via the Web. CDs with SWP were made and given to those who did not live in the dormitories.

The author prepared Web pages to go with the course. These pages consist of frames with appropriate buttons. The starting page consists of a buttonbar on the left, a title bar on the top, and the home page of the author in the middle. The student then clicks on the button related to the course s/he is taking. From there one goes to another set of buttons dealing with various aspects of the course such as a course overview, the grading policy, course notes and exams given in previous years, information about the text, the syllabus, projects, homework assignments, meeting times, etc.

During many of the lectures the instructor had an IBM Thinkpad CDV available for his use. This Thinkpad is so constructed that the back of the screen can be removed and placed on a high intensity overhead projector so that whatever is on the screen of the laptop can be projected for the entire class to see. Students were shown how to access the Web pages, download files, and use SWP. At appropriate times SWP was used in class to solve differential equations, graph the solution to an equation, evaluate an integral, find a derivative, find the first few terms in the series solution of an equation, etc. Students were also shown how to set up the Projects that were assigned.

Three projects were assigned: one on the Web and differential equations, one on the mechanical vibrations of a mass-spring-damping system, and one on Euler's method for solving first order equations. These projects were written in SWP and required the student to use SWP in specific ways in conjunction with employing the standard analytic tools taught in class.

The Second Semester Math Course

The second semester sophomore mathematics course at Stevens deals with eigenvalue problems, Fourier series and separation of variables for partial differential equations, matrices and determinants, multiple integration, surface integrals, and the theorems of Green, Stokes, and Gauss.

In early January the author became a beta tester for Scientific Notebook (SNB), so students were encouraged to use SNB in place of SWP. However, this was not required. About half of the 110 students enrolled in this course did opt to use SNB. Since not all of the students in this second semester course had taken differential equations with the author in the fall, these "new" students all used SNB. SNB is similar to SWP, but it contains a number enhancements and simplifications. It uses the latest version of Maple in its kernel, and allows for connection to the Web. Using SNB one can configure Netscape so that a tex file can be downloaded and opened in SNB directly.

Web pages similar to those described above for the differential equations course were prepared for this course. Thus students were able to get all relevant information regarding the course via these pages. A midi file, which automatically plays music, was embedded in the titlebar as an added "attraction". During the Spring 1997 semester the author's WEB page was accessed more than 4200 times. Clearly students used the WEB pages developed for the course as an integral part of their learning. Three projects were assigned: one dealing with Fourier series, one dealing with matrices, and one dealing with multiple integration.

Wherever appropriate students were encouraged to check the answers they obtained to homework problems using pencil and paper by solving the problems in SNB or SWP. In the past the instructor had assigned the odd numbered problems almost exclusively, since the answers to these are to be found in the text. However, with SNB or SWP the student can easily find the answers to many of the even numbered problems, so these were also assigned.

Performance

Student performance on hourly examinations has been the highest that the author has seen in the more than 10 years that he has been teaching the sophomore mathematics sequence at Stevens. While it is difficult to analyze precisely the reasons for this given that different (but nonetheless similar examinations) are administered each year, the fact still remains that the students appear to have mastered the material better this year when SNB/SWP were incorporated into the teaching/learning experience than in earlier years when this software was not available to them. A contributing factor may be that students can now concentrate on understanding the mathematics and leave the "drudge work" to the software.

Conclusions

There is no question that the use of a program such as Scientific Notebook or Scientific Workplace in traditional mathematics courses at Stevens added new and important dimensions to these courses. This use of computer technology tends to add a more "participatory" dimension to learning the mathematics that is lacking when one uses the traditional mode of instruction. Student evaluations and discussions indicate that many students felt that the experience was interesting and valuable.

The question of the balance between the use of software such as SNB and the teaching of standard techniques is difficult to deal with. This author is opposed to the elimination of the teaching of all pencil and paper activities that can be done with software. On the other hand, having something like SNB available encourages one to think about how the presentation of material should be changed, what material should be eliminated and what should be added. The right mix of computer activities and pencil and paper activities is something that will certainly evolve over time as software develops. Our challenge is to incorporate the new without doing away with the key benefits of the old.

The Walls Come Tumbling Down
Pedagogical Issues and the Web

James Lockard
LEPS (Instructional Technology)
Northern Illinois University
DeKalb, Illinois 60115
e-mail: jlockard@niu.edu

Ana Donaldson
LEPS (Instructional Technology)
Northern Illinois University
DeKalb, Illinois 60115
e-mail: ananiu@niu.edu

Marybeth Koos
School of Art
Northern Illinois University
DeKalb, Illinois 60115
e-mail: marybethk@niu.edu

Our nation is approaching the year 2000 with an education system that is based on the pedagogical methods of the previous century. Many educators are using yesterday's techniques to teach students who are already part of tomorrow. As one writer eloquently stated, "We have allowed our schools to remain in the past, while our children have been born to the future" [Strommen and Lincoln 1993]. A major paradigm shift is under way from *teacher* as giver of information to *educator* as facilitator of student learning (and as a fellow learner) [Downs et al. 1995].

Many authors have described advantages of implementing a constructivist environment in conjunction with the integration of technology into the classroom [Dwyer et al. 1990a, Faison 1996, White 1995, Strommen and Lincoln 1993]. "The constructivist view of learning asserts that learners 'construct' their own meaning/knowledge from the information they acquire. This differs from the traditional view, which assumes a teacher can 'deliver' knowledge to a learner [Dwyer et al. 1990a]. This learning process redirects the emphasis, away from the teacher and toward the student, who must assume increased active responsibility for learning. "The use of the new technologies will have a profound effect on schools. The very relationship between students and teachers will be challenged because the technologies enable learners to gain control of their learning. In the past, schools have been places where people in authority decided what would be taught (and possibly learned), at what age, and in what sequence. They also decided what would not be taught – what would not be approved knowledge. The new technologies provide students access to information that was once under the control of teachers" [Mehlinger 1996].

Educational reform and change must include changing teachers' beliefs and practices. A constructivist student-centered learning environment is characterized by engaged students working as groups with teachers assuming the role of facilitators. Classroom noise and movement conflicts with many traditional teachers' beliefs in the sanctity of classroom quiet and order [Dwyer et al. 1990b]. "Placing emphasis on control, objectivity, managing facts, testing, technology, behavior, and grading (without the corresponding development of the affective, psychological, and spiritual) disconnects, trivializes, and deadens the learning process. We recognize a great learner (and a great teacher) as one who is enlivened, exploring, seeking growth and appropriate challenge rather than compliance and sameness" [Peterson and Hart, 1997]. This is the spirit that must be introduced into the classroom and the nation's teacher preparation programs.

The World Wide Web has the potential to pull down classroom walls and open the world to the student as the learning experience becomes truly student-centered rather than instructor-driven. Based on this concept, the panel members presented pedagogical views and changes that have evolved in their own teaching practices. Key points include the following:

- Many educators see the Web in the limited capacity of a modern day *alternative encyclopedia* without realizing its potential for impacting the classroom experience.

- Rapid Web growth in a short period of time has led to many Web sites that are only a repackaging of old methods in a new media format rather than capitalizing on new capabilities. This limits potential value.

- The Web must be an impetus for pedagogical change.

    - The instructors' role becomes facilitation, not a "talking head."
    - Control of learning passes to the student.
    - Power is rescinded from the instructor.

- Web-based instruction allows students to build on their existing knowledge base and explore creative alternatives to learning.

- The Web supports newer pedagogical perspectives, including constructivism, engaged learning, alternative assessment, and multiple intelligences.

At Northern Illinois University an undergraduate art course and a graduate level education course both illustrate a constructivist view of technology, student-centered learning and the educator in action. In the art class students use the Web as a tool to broaden their artistic experiences. In the education course teachers (as students) explore the possibilities of the Internet and develop their own web sites for classroom use. Both courses represent a major departure from traditional university teaching. The education class web site may be accessed at: http://www.cedu.niu.edu/leps/faculty/donaldson/leit590.

References

[Downs et al. 1995] Downs, E., Clark, K., & Bennett, J. (1995). New Directions for Teacher Education in the Information Technology Age. [Online] Available: http://www.coe.uh.edu/insite/elec_pub/html1995/0810.htm

[Dwyer et al. 1990a] Dwyer, D. C., Ringstaff, C., & Sandholtz, J. H. (1990a). Teacher Beliefs and Practices Part I: Patterns of Change. *(Apple Classrooms of Tomorrow Research Report Number 8)*.

[Dwyer et al 1990b] Dwyer, D. C., Ringstaff, C., & Sandholtz, J. H. (1990b). Teacher Beliefs and Practices Part II: Support for Change. *(Apple Classrooms of Tomorrow Research Report Number 9)*.

[Faison 1996] Faison, C. L. (1996, September-October). Modeling Instructional Technology Use in Teacher Preparation: Why We Can't Wait. *Educational Technology*, 36 (5), 57-59.

[Mehlinger 1996] Mehlinger, H. D. (1996, February). School Reform in the Information Age. *Phi Delta Kappan*, 400-407.

[Peterson and Hart, 1997] Peterson, T., & Hart, T. (1997, Summer). Pedagogy for the Soul: Re-Visioning of Spirit in Education. *Educational Horizons*, 74 (4), 187-191.

[Strommen and Lincoln 1993] Strommen, E. F., & Lincoln, B. (1993). Constructivism, Technology, and the Future of Classroom Learning. [Online] Available: http://www.ilt.columbia.edu/ilt/papers/construct.htm

[White 1995] White, C. (1995). The Place for Technology in a Constructivist Teacher Education Program. [Online] Available: http://www.coe.uh.edu/insite/elec_pub/html1995/0821.htm

Agora: Shaping Social Space from Information Space

Byron Long and Ronald Baecker
Dept. of Computer Science and Knowledge Media Design Institute
University of Toronto
10 King's College Road
Toronto ON Canada M5S 1A4
E-mail: byron@dgp.toronto.edu, rmb@dgp.toronto.edu

Introduction

As an individual browses the web, they encounter others only as authors of the information they browse. Although many people may be reading the same information, they are completely unaware of one another's presence. Rather than communication between peers, the model supported by most interactions on the web is that of presentation and feedback, where there is a distinct difference between the status of the participants.

A tool supporting a more equitable basis for communication between individuals seems to be required. Such a tool should be able to give an awareness of others while browsing — showing others accessing the same content, perhaps indicating a commonality of interest. Further, in order to capitalize upon this awareness, it is necessary to allow communication between these people so that they might discuss their shared interest and perhaps forge longer-term relationships. Through these relationships, it is possible that communities will form, centered on a particular topic or a specific meeting place.

For such a community to be stable, it must be possible to leave persistent artifacts — to relay some of the history of the community. This requires the addition of asynchronous communication tools. Considering the relatively small amount of time people are likely to spend at any one information resource, the ability to communicate asynchronously is considerably more important than it might be in a community where its members are more likely to be together. Real-time communication tools allow the members of a community to meet one another and form the community itself. Asynchronous tools are required to maintain the community in the long term, and for the initiation of new members.

Agora [Long 97] is designed to fulfill these needs. It provides both real-time and asynchronous communication within the information pages of the World Wide Web. Designed to assist formation of Internet communities, it provides the ability to determine who else is browsing an information space of interest, to communicate with them in real-time, to view who has recently come and gone, to read and post messages of interest to the community, and to send and receive personal messages from others in the community.

There are other tools, such as *Virtual Places* and *WebTalk* [Donath & Robertson 94], that are designed to allow communication with others in a web page. These tools provide only real-time communication, however, and do not have a "dialog history" [Long & Baecker 97], a representation of recent communication that allows a new participant to acquaint themselves with recent current discussion.

Functionality

The primary purpose of the Agora client is to support identification of other visitors to a web page and allow real-time communication between them. Currently, this communication is text-based, with each participant being able to send phrases to either everyone present, a small group of people, or an individual. In addition to text phrases, it is possible to 'perform' actions by having a description of the action relayed to the other participants.

There are also several functions that support a sense of community history within the system, including a list of recent visitors, a 'bulletin board' for the posting of news, and private email boxes. The list of recent visitors provides a limited history of the participants in the community, and provides a way to send an email message to or examine the profile of a user who is not currently present. This offers an advantage over most real-time communication systems with respect to coordinating a meeting within the system. Most systems give no indication as to whether another user has already left or has yet to arrive, and do not allow messages to be sent to users who are offline.

The bulletin board allows messages of general interest to be posted and for long-term open discussions to take place. The news and discussions may augment the content of the web page the community is attached to or may be relevant to the community itself. In either case it enhances the perception of community stability — it shows that there have been interested community members for some length of time.

Finally, the ability to send and receive email allows relationships with other individuals to be pursued without requiring constant coordination of real-time meetings. By default, such mail messages are received when the system they originated from is re-visited, but the ability exists to have the mail routed directly to a user's email address. It is possible to send on the mail and handle replies without giving either party the actual email address of the other. By having both parties send their mail through the Agora server, it is not be necessary exchange addresses. This allows for some real-world anonymity despite active participation within an Agora community.

Interface

The Agora client is designed as a Java applet that can be inserted into any web page. Because of the small physical size of the applet, and because spawning a large number of windows would cause the browser to be obscured, it was important to allow all the major features of the client to be visible in a limited space. The user is initially presented with a login screen where they can identify themselves as a return visitor or create a new account.

Once the user logs in, the current thread of real-time conversation is displayed, along with a field for the user to type phrases to be sent to others. In addition, there are controls for sending actions and for sending a message to a restricted set of people. In the remainder of the window is an area that can be set to display a range of information, defaulting to a listing of the current visitors to the page. The other 'pages' of information include a list of people who have recently visited the site, a list of email messages addressed to the user, a listing of the 'bulletin board' articles, a display of the current message or news article, and a profile of the currently selected user. The paged design allows access to many functions within a small space, while the constant presence of the real-time conversation window allows a user to join the ongoing conversation at any time without interrupting the task they are currently engaged in.

Architecture

Agora is implemented as a client-server system. The Agora client, designed to run within a web browser, communicates with the Agora server, which runs on the web server hosting the page on which the client resides. The server then propagates any relevant messages to the clients currently connected to the server. The server is also implemented in Java. It manages the mail and news for the pages supporting the client. In addition, it manages the real-time communication channels for each user and maintains the user profile.

A single server can support a number of 'groups'. Each group is a separate communication space, sharing only the profile information for each user. The client connects to only one such group, although clients on different pages can each connect to the same group.

Conclusion

An initial version of Agora has been completed and has undergone a pre-test as a supplemental channel of student communication for a graduate class. It is currently being used on the home pages of the Knowledge Media Design Institute as part of a study of the system's effectiveness. The results of this study will be discussed at the conference.

Supporting a community inside the bounds of what is primarily a broadcast information space enhances the experience of those who use it. In supporting conversation with other navigators of the information space, whether for social reasons or to discuss the information at hand, human expertise for social interaction can be exploited to make the information space more salient and more enjoyable.

References

[Donath & Robertson 94] Donath, J. and Robertson, N. (1994). The Sociable Web. *Proceedings of the Second International WWW Conference*. Chicago, IL. http://judith.www.media.mit.edu/SocialWeb/SociableWeb.html

[Long 97] Long, B. (1997) Shaping Social Space from Information Space. Master's Thesis in progress, Department of Computer Science, University of Toronto.

[Long & Baecker 97] Long, B. and Baecker, R. (1997). A Taxonomy of Internet Communications Tools. *Proceedings of WebNet '97*. Association for the Advancement of Computing in Education, Charlottesville, VA.
http://www.dgp.utoronto.ca/people/byron/webnet/taxonomy.html

A Framework for Developing Web-Based Educational Applications: Using a Relational Database to Connect the Design of Authoring and Navigational Tools

Yibiao Lu, Matthew J. Koehler, Richard Lehrer
Wisconsin Center for Education Research, University of Wisconsin-Madison
yibiaolu@facstaff.wisc.edu, mkoehler@students.wisc.edu, lehrer@macc.wisc.edu

Wei-Tek Tsai (tsai@cs.umn.edu)
Department of Computer Science, University of Minnesota, Minneapolis

The World Wide Web (WWW) uses the non-linear format of hypertext to provide readers point and click access to networked multimedia information (text, graphics, audio, video, etc.). This new format for information has created new forms of literacy for readers and writers alike. However, this format has also raised some concerns for researchers interested in using the WWW for educational purposes. For readers who are accustomed to the conventions of traditional, linear texts, the organization of hypermedia documents can be obscure and confusing. Additionally, unlike readers of traditional texts, these readers are now required to sequence (or navigate) the information in this non-linear space. Faced with an unfamiliar representation and unfamiliar navigating tasks, many readers have become confused. This confusion has been popularly termed "lost in hyperspace" [Castelli, Colazzo & Molinari 1996].

The hypermedia format introduces problems for authors as well. While there are familiar tools for writing traditional, linear texts (i.e., word-processors), the tools needed for writing in hypermedia are less obvious. Hypermedia authors require tools that can manipulate each of the supported media formats (text, video, audio, graphics, simulations, etc.). Not only do authors need tools to create and edit pieces of multimedia information, they also need facilities to link these pieces of information. Furthermore, authors need tools that make these documents easy to revise and maintain. For example, authors should not have to update every link to a given piece of information if something in that piece of information changes (its location, its name, etc.).

Both authors and readers alike face problems in managing and working with non-linearity. However, these issues are often treated as separate problems as researchers focus on only one of these two related problems. For example, Bevirt [Bevirt 1996] described a variety of navigation tools such as home page links, reference table links, "previous" and "next" links, table of contents, and search engines as important tools for understanding the structure of the information. In addressing the problems that authors face, a variety of hypermedia tools have been developed, including GETMAS [Wong, Chan, Cheng & Penh 1996], HM-Card [Mayrhofer, Scherbakov & Andrews 1996], Hypercourseware [Siviter & Brown 1992], and HCC (HTML Course Creator) [Carver & Ray, 1996] and so forth. There is very little work, however, that addresses these two problems in conjunction with one another.

The problems that readers and writers face are in fact quite similar. Authoring requires tools to create, manage, visualize, and modify the non-linear structure of hypermedia documents. Readers require tools that convey the non-linear structure of hypermedia documents and help them to navigate around this structure. In short, the problems that authors and readers face with hypermedia documents are the same: Both need tools to support their interactions with the structure of the hypermedia documents. This paper tries to address these two problem jointly, based upon on a model we propose here.

We contend these two problems are best addressed by employing a common representational structure to be used by authors and readers alike. If the design of authoring and navigation tools are based on different structures, then creating and maintaining a system becomes a challenging task. For example, if these two structures are treated differently, changes in the authoring structure requires additional work to reflect these changes in the structure (and navigational tools) presented to readers. To provide this common structure, we propose that relational databases are ideally suited to provide this common foundation. Relational databases have long been used in information systems and are capable of storing large amounts of data in a highly structured, consistent way. Relational databases are also adept at representing relationships between different entities within a database, thus providing the foundation for representing hypermedia documents. Using a relational database to represent the structure of an information system, authoring becomes a process of using tools to transparently create, modify and refine the database structure. Likewise, reading becomes a process of using the navigational tools to transparently explore and traverse the database structure.

Using this approach, we propose a model for design and developing a web-based application which consists of four steps.

- First, system designers must create an organizational scheme for the information system. This structure identifies the semantic types of information, the media form for information, and the link types needed to connect the various pieces of information.

- Second, the navigation and authoring tools that are needed to support this organizational scheme must be identified.

- Third, a database must be designed that is capable of supporting the organizational scheme as well as the navigational and authoring tools.

- Fourth, the authoring and navigational tools are implemented and integrated with the underlying relational database.

References

[Bevirt 1996] Bevirt, B.: Designing Hypertext Navigation Tools. *In Proceedings of WebNet 96: World Conference of Web Society*. San Francisco, 1996. 1-6

[Carver & Ray 1996] Caver, C.A. and Ray, C.: Automating Hypermedia Creation and Maintenance. In *Proceedings of WebNet 96: World Conference of Web Society*. San Francisco, 1996. 82-87.

[Castelli, Colazzo & Molinari, 1996] Castelli C., Colazzo L. and Molinari A. (1996). Getting Lost In Hyperspace: Lessons Learned and Future Directions. *Proceedings of ED-MEDIA 96: World Conference on Educational Multimedia and Hypermedia*. Boston, 1996. 124-130..

[Mayrhofer, Scherbakov & Andrews 1996] Mayrhofer V., Scherbakov N., and Andrews K.: HM-Card: A new Approach to Courseware Production. *Proceedings of ED-MEDIA 96: World Conference on Educational Multimedia and Hypermedia*. Boston, 1996. 433-438.

[Wong, Chan, Cheng & Penh 1996] Wong W.K., Chan T.W., Cheng Y.S., and Penh S. S.: A Multimedia Authoring System for Building Intelligent Systems. *Proceedings of ED-MEDIA 96: World Conference on Educational Multimedia and Hypermedia*. Boston, 1996. 708-713.

CoProcess: A Java-based Environment for Collaborative Process Management Over the Web

K. J. Maly, H. Abdel-Wahab, R. Mukkamala, A. Gupta,
M. Kholief, S. Dittakavi
Department of Computer Science
Old Dominion University
Norfolk, VA 23529-0162
maly@cs.odu.edu

C.S. Vemuru and H. Syed
Innovative Aerodynamic Technologies
Poquoson, VA 23662

Introduction

CoProcess is a Java-based environment to facilitate collaborative process management over the World-wide Web (WWW). It can support a wide variety of applications that may involve filling forms, sending e-mail, searching a database (or web) for information, establishing a teleconference within a group, collaborative writing, reviewing, and editing of documents, and other video and audio record/playback facilities. In this paper, we describe one such application developed in this environment---a collaborative tool to implement all processes involved in writing a graduate thesis. *Thesis WebBook* guides the students through the entire process providing all the necessary forms, establishing necessary databases for work-in-progress, aiding the selection of topics and advisor, and facilitating communication between the student and the committee members.

Background

While the recent development in web technologies and the need for tools to aid collaboration among groups of people has encouraged us to start developing CoProcess, the primary impetus comes from two recently developed concepts: Electronic handbook and CoReview (Maly et al 1995).

The Electronic Handbook (EHB) concept, introduced by Dr. Barry Jacobs (NASA Goddard Space Flight Center), automates any process involving documents in the broadest sense. For example, an organization's business processes would be ideal candidates for an EHB. The concept is to write a particular process, e.g., purchasing, as a chapter in an organization's electronic book (or manual). CoReview, an interactive document exchange and review tool, is developed by Innovative Aerodynamic Technologies (IAT) and the Old Dominion University. It provides the ability for a group of geographically distributed users to work together as a team without the need to travel for face-to-face meetings. It addresses the need to evaluate proposals over local and wide-area communication networks. In our work, we have combined the EHB concept and the facilities offered by CoReview with the portability and security features offered by Java (Deitel and Deitel 1997) to design CoProcess.

CoProcess

As stated earlier, CoProcess is an environment to effectively and efficiently implement processes involving collaborative efforts of several individuals who may not be located at a single site. In this paper, we illustrate the CoProcess concept using *Thesis WebBook*, a tool to aid graduate students in following the steps involved in conducting research. The following sections describe the functionality of Thesis WebBook and briefly summarize the implementation details.

Thesis Webbook

The Thesis Webbook is a tool used to develop the various features of CoProcess to support the process-oriented Webbook. This tool enables students and faculty, working on a project such as a Master's or Ph.D. thesis, to resolve questions or exchange ideas, to follow proper procedures, and to keep track of the student's progress.

The tool provides for role-dependent viewing which enables the users to assume different roles. The two roles envisaged are Faculty (Advisor and Committee Members) and Student.

Each role has a different view of the Webbook. The student's view of the Webbook has four main chapters: Start the Thesis, Thesis Proposal, Thesis Defense, and Acceptance of Thesis. Faculty's view enables them to browse any of the thesis documents of their students, browse the list of Committee members, start a collaborative session with some or all of the Committee members, annotate any of the documents, and to query the status of a student's thesis.

The Thesis Webbook also provides for general purpose tools like text editors, e-mail and collaborative sessions which would be used by students throughout the process. The collaborative session tool allows the students to establish communication with their Advisor and Committee members. It also provides for additional features like multi-party chat, audio-chat, distributed annotations etc. the details of which are discussed in the following subsection.

Implementation

The Thesis WebBook is implemented using the tools provided by CoReview and the facilities provided by Java. The two main issues that arose due to the use of Java were security and communication. The current version of *Thesis Web-book* (see http://www.cs.odu.edu/~iat/webbook) includes the global and local Daemon implementation as well as the setup and invite features of CoReview. A global Daemon, which runs on the WWW server of each network involved, could be the source of communication from any participant in a CoReview session. In addition, each participant has to run a local Daemon on his machine before he can use any feature of CoReview. The global Daemon communicates with other global Daemons, local Daemons, a session controller and Applets' connections. Each of the CoReview's features is implemented as a Java Applet and thus can be run on any platform. Thus, an organization such as a university can buy a site license which would set up, when installed on appropriate networks, the global Daemons. Individual students and/or faculty members would buy an individual copy of the software which, once installed on that student's workstation, will run the local Daemon.

Future Plans

We are currently in the process of supporting audio and video features to Thesis WebBook. The new features will enable the advisors or committee members to leave their comments or concerns about the thesis or research work in an audio/video file when it is not possible for all parties to meet together. We are also in the process of adding the annotations feature to the Thesis WebBook. In the long term, we intend to develop a platform independent tool library which could be molded to support any process-oriented application. The current notion of a Thesis Webbook is being used to realize the features needed to be incorporated in CoProcess.

Acknowledgements

This work is supported in part by NASA GSFC under the STTR-96 program (Contract NAS5-33235).

Literature References

[Deitel and Deitel 1997] Deitel, H.M., & Deitel, P.J. (1997). *Java How to Program.* Prentice Hall Inc.

[Maly et al 1995] Maly, K.J., Abdel-Wahab, H., Mukkamala, R., Gupta, A., Prabhu, A., Syed, H. & Vemuru, C. S. (1995). Mosaic + XTV = CoReview, *Computer Networks and ISDN Systems*, 27, 849-860.

A CAL PACKAGE for HTML BASICS

Dr Moira J. McAlister
School of Computing & Information Systems
University of Sunderland
St Peters
St Peters Way
Sunderland
Tyne & Wear
England, UK
SR6 0DD
email: moira.mcalister@sunderland.ac.uk

Introduction

The School of Computing and Information Systems has recently been located in a new building with an open plan architecture. Computer equipment is subsequently structured into zones, each zone containing a bank of Macs, PCs or Sun Workstations. This format of design thus requires the lecturer to adapt their mode of delivery. So new innovative methods of instruction must be used to successfully address a potentially dispersed audience. One module which is adopting this approach is Information Products currently taught on the second year of BSc Technology Management.

Currently, students work through handbooks in hands-on laboratory sessions, it is a time consuming process which often leads to rote learning as opposed to retention. Therefore, a more interactive approach was devised allowing the students to utilize a CAL package which involved them developing the code for HTML documents and viewing the results directly on the screen.

The development of such courseware has four basic strands; [Rowntree,1982]: identify course aims and objectives, develop necessary learning experiences, evaluate the effectiveness of learning experiences and improve the experiences in light of the evaluation.

Previous papers have addressed a variety of issues related to the development of educationally sound courseware; [Culwin and Marshall,1996];[Marshall and Hurley,1996]; [McAlister and Grey,1996]. User interaction [Gagne et. al. 1992]; [Schank,1993] being viewed as being of prime importance in gaining and maintaining attention and providing stimulating material and encouragement via feedback to the student. A variety of hypermedia tools have been developed such as GETMAS©, HMCARD©, Hypercourseware©, Hypertactics©, ISSAC©, MALL©, Metaplant©, Nestor© and those developed at academic institutions [McAlister and Smith,1997]. A variety of public domain, shareware, and commercial software tools have been distributed for the development of HTML documents such as HTML Writer©, HTML Assistant©, HotDog©, HotMetal© and Internet Assistant©. None of these tools provide support for libraries, point-and-click and no HTML knowledge.

Design

The system under development, which is currently at the first-cut prototype stage, uses Visual Basic to emulate a HTML editor and a WWW environment. Although a number of editors are available, some of which provide templates such as HotDog©, they still require a knowledge of commands and menus for Web page design. The first-cut prototype provides the students with tutorial exercises in five basic aspects of writing a Web document, namely: HTML document structure, Text formatting, Linking/images, Forms, Frames. These were considered to be the basic building blocks of any Web page. The style and complexity with which a user can manipulate text and graphics coming directly from the design principles used to teach basic skills. A storyboarding technique was taken to screen design which involves the student being provided with a statement of the problem and a description of the tags which could be used to answer the problem. The environment allows the student to position a tag and set attributes in the HTML document structure where it is translated into the appropriate HTML code. The WWW screen then displays the result which could be correct or an error. The student is given three opportunities to select the correct tag after which the system carries out the process automatically. As the student makes successive attempts to select the appropriate tag more detailed help is provided via dialog boxes.

Implementation

An initial system has been designed in Visual Basic 3 and Access; [Sparrow,1996]. The main engine of the system is a database. The database is made up of several tables. There is a table for each HTML object such as a heading, image and so on. Each one of the tables has its own specific attribute related to that object such as align, underline or width, forming a template for each object. A contents table was then added to the database showing the order in which objects should be placed on Web pages. The user can then step through the contents of each table and alter attributes for the relevant object. When the desired changes have been made the system then generates HTML code from the information in the tables. A browser can then be run to view the results.

The prototype produced was successful in determining the feasibility of generating code and passing that code directly to a browser such as Netscape©. However some points to note include that the prototype produced worked on limited HTML objects specifically those commands which are most commonly used in HTML, the use of templates somewhat restricts the flexibility of the code produced which although may be viewed cannot be altered and the use of Java as the application language could enhance the system by permitting the use of a full spilt screen facility since the ability to write such a piece of code is dependent on both applications, the editor and the viewer, being capable of operating in a Netscape browser environment.

Further Work

The use of an application language such as Java allows the prestored templates written in VB to be converted to dynamic library routines which can be updated, added to or deleted as the tool expands and editors adjust to new requirements. The interactive nature of Java applications permits tests to be developed which can be linked or focused on a particular subject area, such as text formatting. It is envisaged that the tests could be generated as modifications to the library routines thereby self-generating a series of questions on a particular subject area.

References

[Culwin and Marshall,1996] Culwin F. and Marshall A. (1996) The Design and Provision of Software Engineering Education Over the Web, , Proceedings of the 5th International WWW Conference '96, pp 217-226, Paris, France, May, 1996.

[Gagne et. al. 1992] Gagne R. Briggs L. And Walter W. (1992) Principles of Instructional Design, Harcourt, Brace, Jovanovich.

[Marshall and Hurley, 1996] Marshall A. and Hurley S. (1996) The Design, Development and Evaluation of Hypermedia Courseware for the World Wide Web, , Journal of Multimedia and Its Applications, Vol. 3, No. 1, June.

[McAlister and Grey,1996] McAlister M. And Grey M. (1996) Intelligent Courseware for the WWW, , 4th Annual Conference on the Teaching of Computing, Providing Quality Computer Science Education, Dublin, Ireland, August, 1996.

[McAlister and Smith, 1997] McAlister M. And Smith P. (1997) Using Undergraduate Projects to Develop Tools, Techniques and Courseware for the WWW, CAL '97, Exeter, March.

[Rowntree,1982] Rowntree D. (1982) Educational Technology in Curriculum Development, , Paul Chapman.

[Schank,1993] Schank R. (1993) Learning via Multimedia Computers, Communications of the ACM, Vol. 36, No. 5, pp 54-56.

[Sparrow, 1996] Sparrow, G. (1996) HTML Generator, BScIT Final Year Project.

Issues: The Internet and Society

Leah P. McCoy
Department of Education
Wake Forest University, USA
mccoy@wfu.edu

As we rush forward as fast as technology will take us, we need to step back and consider the impact of these vast technological changes on our society. This paper describes the results of a university seminar where students and faculty critically reviewed the Internet and its impact on our society. Identified issues have been grouped into seven broad categories: regulation, privacy, education, commerce, communication, entertainment, and addiction. A brief description and questions for further discussion are given below for each area.

Regulation

The first issue, regulation, is perhaps the most controversial. The U.S. Supreme Court recently overturned the Communications Decency Act, calling it a violation of First Amendment Rights [CEIC 1997]. This leaves Web materials unregulated. One must consider that the Web is increasingly used by young children, and that it is also a major vehicle for delivery of pornographic pictures. It is difficult to keep these two extremes of purpose separated. The "formula" for making a bomb similar to the one used in Oklahoma City is easily accessible online. Clearly, there are both valuable and also unacceptable materials on the Internet, but the problem is to determine whose values will define appropriate and inappropriate.
(Is the Web an open forum, protected by the First Amendment? Is pornography inevitable? Can children be shielded? Should the Web be regulated?)

Privacy

Another important issue is privacy. While the Internet can help us to perform a number of tasks, we often are taking a chance when we transmit confidential information. Security is violated and crime may be promoted if bank information, credit card information, or medical information fall into the wrong hands [Stoll 1995]. Newly developed encryption software may be a viable solution to this problem.
(Would you feel comfortable managing your bank account online? Is giving your credit card number online different from giving it over the phone? Can encryption software provide adequate security?)

Education

The impact of the Internet on education at all levels is tremendous. Use of the Web is making education richer and more accessible for students and teachers at all levels. Use of the Web for research has changed the concept of the library from kindergarten through college [Negroponte 1995]. The Web is an important source of interactive information and up-to-date news. There is some concern that there is disparity in computer access for students of different economic levels in different schools. Educators at all levels are modifying teaching practices to include the Internet [McGuffey Project 1997].
(Is use of the Internet changing instructional practices in K-12 and college classes? Should the government provide leadership and financial support for less-affluent school systems as they implement computers?)

Commerce

The Web is becoming an integral part of business, and a convenient mode of shopping for consumers [Gates 1995]. Advertising, in all its annoying forms, is apparent on many webpages. Commercial sites are used for advertising for products and services to be purchased both online and offline. *(Will the Internet become a big online catalog? Should online advertising be banned?)*

Communication

Communication on the Internet is an important issue. Email is the most popular facet of Internet use, closely followed by the World Wide Web [Negroponte 1995]. Recent developments that allow email attachments of pictures and video, and even real-time audio and video, are even more attractive. The Internet has been used successfully to facilitate communication for the physically, cognitively and emotionally handicapped [Lindsey 1993].
(What kinds of adaptive devices are used by handicapped persons? Is email efficient and effective?)

Entertainment

Entertainment materials on the Web are very popular. A recent survey of Web users found that favorite sites included sports, movies, music, and celebrities [Heichler 1997]. Multimedia computer systems are very appealing and allow users to interact in visual, audio, and video modes.
(How many Grammy winners have homepages? Are Web-based movie reviews useful? Do webpages have the same range of credibility as print media?)

Addiction

There is evidence that some users are addicted to the Net. They spend large amounts of time interacting with the computer in anonymity and isolation [Stoll 1995]. The Net may produce a false sense of community with limited actual personal contact and social life.
(Is there a danger to persons who spend too much time on the Net?)

It is imperative that we continue to scrutinize the effects of widespread Internet use on our society. Discussions by concerned and knowledgeable users must continue in many different forums.

References

[CEIC 1997] Citizens Internet Empowerment Coalition (CEIC). (1997). Supreme Court Rules CDA Unconstitutional [On-line] Available: http://www.ciec.org/ [1997, August 4]

[Gates 1995] Gates, W. H. (1995). *The road ahead.* New York: Viking.

[Heichler 1997] Heichler, E. (1997, May 12). Internet lacks content for women. *Computerworld*, p. 64.

[Lindsey 1993] Lindsey, J.D. (Ed). (1993). *Computers for exceptional individuals. Second edition.* Austin, TX: Pro-Ed.

[McGuffey Project 1997] McGuffey Project (1997). 21st Century Teachers Network. [On-line] Available: http://www.21ct.org/index.html [1997, August 4]

[Negroponte 1995] Negroponte, N.(1995) *Being digital.* New York: Vintage Books.

[Stoll 1995] Stoll, C. (1995). *Silicon Snake Oil.* New York: Doubleday.

An Authoring Tool for Constructing Interactive Exercises based on Automatic Generation of Client-side Functionality

Sean McKeever, Damien McKeever, John Elder.

Department of Computer Science,
The Queens Univerity of Belfast,
Belfast, BT7 1NN.

s.mckeever@qub.ac.uk

Introduction

Dynamic client-side interactivity can be harnessed, and tools developed to automate the construction of interactive web documents or website 'components'. Electronic catalogues, multimedia presentations, document navigation components, and in this case, interactive exercises are used as exemplars. Many useful applications of interactive WWW documents relate to areas best addressed by those who have expertise in the relevant area but without technical internet development skills. Simple authoring tools address this need. The tools should facilitate top down design and development, focusing on the semantics of the task rather than the physical implementation mechanisms. The tools should also separate the semantics from the presentation characteristics, such as visual layout and styling, while allowing the author considerable stylistic freedom.

Server-side CGI methods have been used to implement similar functionality [Kuntz and Walthall 96]. The client-based approach offers improved responsiveness to user actions. Server independence frees the developer from the overheads of server-based implementations. This is especially suitable where the site is hosted by a third party such as an internet service provider (ISP). Off-line distribution and usage is possible, effectively using the web browser as a cross-platform player application. Browser support for HTML features such as frames and inline scripts is required.

A component consists of a structured data description, a Javascript runtime engine, a set of parameters specifying functional and presentation characteristics of the component instance, and a HTML-based graphical user interface. The runtime engine is responsible for extracting, processing and presenting the required data, according to predefined layout conventions or parameters, in response to interface actions. Any interactivity associated with the data itself requires appropriate event handlers and calls to the engine to be incorporated.

Interactive Web-based Exercises

Interactive exercises offer goal-oriented learning and are widely used in computer-based training. They can be most effective when combined with performance measurement for self-testing, and feedback incorporating learning recommendations and active references to supporting course material.

A general model of an exercise has been developed. A small database can be constructed for an exercise instance. This database is converted into a HTML Form with embedded event handlers, either statically (by the authoring tool) or dynamically (at runtime by a Javascript engine).

At runtime when the value of a form element is changed, the Javascript engine is called and an identifier for the target element is passed. The Javascript engine extracts the details for the question, compares the answer given with the expected response, and updates a stored user performance table. The result can instantly be displayed to the user, along with a link to any question feedback. The performance can optionally be displayed as a running score, and/or at the end of the exercise, in conjunction with feedback and recommendations.

Structure of an exercise

An exercise is defined as follows:

Exercise = Context, {Question}.

Context = Exercise title, Author, Email, Home page, Logo, Header, Footer.

Question = Title, Task, Response [, Feedback, Attempts].

Task, Feedback = text | HTML.

Response = ResponseType, ResponseValues, ResponseExpected.

ResponseType = text | radioGroup | popupMenu | selectableList | imageSet | checkboxGroup| multipleSelectionList.

ResponseValues = {String}.

ResponseExpected = {String}.

Attempts = positive integer.

Thus an exercise is a sequence of questions in a given context setting. A question consists of a title, a task specification and a response. A response can take the form of text entry, multiple choice (radiobuttons, popup menu, selectable list, clickable images), and multiple selection from a set of options (check boxes, multiple selection list). To permit automatic checking of the response, a set of expected responses must be specified. User feedback and performance-related recommendations may be added. Questions and feedback can be text or HTML, therefore these sections can contain links to related course material, or embedded multimedia objects, such as video, animation or sound. The number of permissible attempts may be specified. A timing element may also be introduced.

The Exercise Authoring Tool

The authoring tool provides a point and click GUI for constructing the semantic datastructure. The data is maintained in a file for future editing and modification. The authoring process involves creating questions using the menu or toolbar commands, and selecting or entering the required structured information. The question and feedback data is entered in the allotted text fields (text or HTML is supported). A HTML file authored using any web page layout application or HTML editor can be imported. Alternatively more limited mark-up can be carried out using the tagging facilities provided by the authoring tool. The response type is chosen from a predefined set and a list of possible response values can be entered. The runtime representation of the response type is displayed. The correct answers are selected or entered. Questions may be renamed, cut, copied, and pasted for reordering or reuse in other exercises. The current runtime version may be viewed, a working version of the component is generated and the browser defined in the users preferences is launched.

Alternative functional, stylistic and contextual features may be graphically enabled, disabled or set. For example the layout variants and their associated functionality may be selected by clicking on a toggle button for each functional requirement. Stylistic characteristics for each frame may be modified using the controls provided. For example, colour settings can be typed in, selected from a pull-down menu, or from a colour wheel. The attributes as a whole can then be poured into each frame individually using the paint pot tool. Exercise title, logo, header, footer, email and home page contact information can be entered. This will be displayed in the runtime document in predefined locations.

Summary and Future Research

The authoring tools offer a higher-level approach to the development of interactive web documents while minimising the requirements for technical knowledge of internet development languages and mechanisms. They enforce a greater separation between content, presentation characteristics and functionality. This allows users to focus on the logic and content of the component rather than on the HTML and Javascript. Incremental development gives a user the flexibility to develop as much as time permits, beginning with a simple outlinedeveloping the contents, and fine-tuning the stylistic or presentation attributes over time.

The tools aim to build on top of and integrate with existing systems, supporting the import of HTML created with other applications. While allowing considerable freedom in determining the structure, behaviour and style of an exercise, this type of software is not possible without enforcing some standard structuring and interface layouts. Added flexibility and customisation options may be added at a later date.

The components require frames and Javascript support. Browsers vary in both their interpretation of the Javascript language and the supported implementation of HTML. A limited variety of browsers can be used to access the full functionality of the resulting exercises.

The future research and development direction is focused on integrating this work with ongoing research on a high-level design notation and description language, known as Hypermedia Design language (HDL). HDL is a concise notation for use in the systematic specification of multimedia or web-based systems. The design of an authoring environment built on top of HDL, with an extensible plug-in architecture, is on-going . This authoring tool is centred around an editable logical tree view of a site domain. A leaf node in the tree is considered to be a simple URL or a web component descriptor. Descriptors can be converted to runtime components consisting of a combination of HTML, Java applets, and Javascript engines. Support of dynamic data access management can be incorporated by embedded applets, for loading data from the server on demand. Additional web component types will be identified and added.

References

[Elder, McKeever & McKeever 1995] Elder John, McKeever Damien and McKeever Sean (1995) Multimedia Publishing Tools for Software Engineering. *IEE Digest* 1995/038.

[Kuntz and Walthall 96] Kuntz David, and Walthall, Ned. The Remote Item Development Environment (RIDE), *Proceedings of Webnet '96*, San Francisco, CA, USA (1996).

Building Context-Sensitivity Through Behavioral Prototyping: The Development of an Internet-Based Mentoring Program for 5th-12th Grade Students

Jay Melican
Institute of Design
Illinois Institute of Technology
Chicago, IL 60616
USA
(jaym@id.iit.edu)

Eli Blevis
Institute of Design
Illinois Institute of Technology
Chicago, IL 60616
USA
(eli@id.iit.edu)

Introduction and Motivations: A National Literacy Initiative

In May of 1995, at Beechtree Elementary School in Falls Church, Virginia, United States Secretary of Education Richard Riley formally launched the American Initiative on Reading and Writing, a federally-sponsored effort to encourage the development of critical literacy skills across all strata of the country's population, particularly in its young students. In that announcement, the Secretary reemphasized the Department of Education's commitment to innovation and its openness to the exploration and full exploitation of developing technologies. Originally funded by the U.S. Department of Education as part of the national literacy initiative, our team at the Institute of Design at Illinois Institute of Technology has been working on the development of Writing Exchange, a program that leverages Internet communications technologies to provide informal, mentor-assisted writing support to school children.

Writing Exchange is a work in progress, and this paper documents the project in its initial stages of development. In addition, it serves as a case study exemplifying the design and development processes and the "user-centered" design philosophy advocated at the Institute of Design. Accurately predicting human behavior and designing products that "work" and "make sense" in specialized social contexts can be extremely difficult. Consequently, at the Institute of Design, we employ a number of user-centered design methods in the development of educational technologies. Using techniques that include user observation and behavioral prototyping testing, we believe we can build critical context-sensitivity into the products we develop.

Concept Prototyping: Exploring User Scenarios to Refine the Project Concept

Our initial project brief called for the development of an elaborate World Wide Web site that could function to provide student mentor/apprentice teams with an extensive information resource and activity center. A series of concept prototypes based in imagined user-scenarios led our design team to focus on behavioral aspects of the exchange of communications between fifth grade writing apprentices and their high school mentors.

Behavioral Prototyping: The Pilot Study

The Writing Exchange pilot study involved two Chicago area schools. A class of fifth grade students and their teacher, Dr. Jane Rosen, at the Newberry Academy along with a group of volunteer, high school mentors under the guidance of the head of the English Department, Lucy Kowalski, at Von Steuben High School put Writing Exchange into action for the first time. Both Chicago Public Schools were selected on the basis of near technological readiness for the project and diversity of student population.

Using off-the-shelf software to prototype a system of communications that included all the functionality initially envisioned for the final product, we were able to get the exchange rolling quickly so that hypotheses about middle- and high school student behaviors could be checked. Both groups of students were using ClarisWorks for word processing. Eudora was introduced as the telecommunication software, and the Internet connection was provided through our university server. After we had given demonstrations of software use at both schools and shared the mysteries of user names and passwords, the pilot study began.

Feedback and Findings

Each member of the design team took responsibility for observing the e-mail interactions of six apprentice-mentor pairs. By the end of the three month pilot test, we realized that we had made several incorrect assumptions about student skill levels and about the context in which the technology was used. This was a time to rethink the overall structure of the writing collaboration. This first phase of behavioral prototyping and user observation was fruitful in helping us consider seemingly simple issues such as: the ratio of students and available in-class time to computer accessibility; time and frequency of use necessary for students to develop an "e-mail culture"; observation time needed to monitor developments in writing; the importance for the fifth graders of the ability to print their e-mail, and much, much more.

Observations made during this initial behavioral prototype test were recorded and discussed with the design team. Possible fixes were suggested and collected as considerations for the subsequent iteration of functional specifications. (See Table 1)

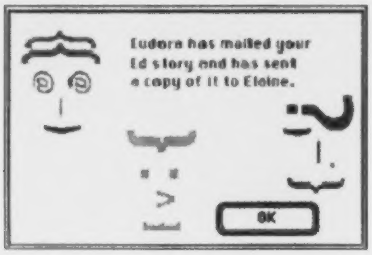
Conclusions

As a result of the observations made during the prototype test, software specifications for automating the mentor/apprentice matching process were developed; features for a special writing/editing application to replace the commercial word processor became more specific; the method for mentor comment was refined, and the need to keep apprentice and mentor "roles" separate was built into the software scheme. (See Figs. 1-4)

Had we skipped this stage of behavioral prototyping and user observations, we would have overestimated students' technical expertise on the computer, dismissed the collaboration issues between the two teachers and their technical support staff, presumed that access to the Internet was sufficient motivation and reward for the high school students' participation -- in other words, we would have extrapolated from our own technical, adult experience and failed to meet the real users in all their enthusiasm, confusion and diversity. Having done the pilot study early, we were able to adjust the project objectives to embody a better fit for the user and to improve the integration of the overall system.

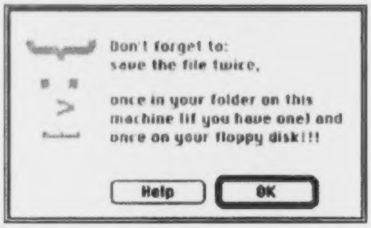
Table 1. A few examples of the many findings and design implications recorded during the initial behavioral prototype test.

<table>
<tr><td colspan="3">Instituting an e-mail culture</td></tr>
<tr><td>Observation/Problem: There was no past experience with an "e-mail culture" in either population. This meant that students were not used to checking for mail each day and that we ran the risk that the students (especially the fifth graders) would become disappointed after not having received mail for several days in a row.</td><td>Insight: One design team member coined the term "e-mail threshold," referring to the number of e-mail messages that a student would need to start receiving on a daily basis in order to feel compelled to check his/her e-mail regularly.</td><td>Possible Fix: We could try packing the students' "in" boxes until they have adopted the habit of checking their mail regularly. We could increase the amount of mail the students receive by signing them up for listservs or by showing them how to sign up themselves. If the content of available listservs proved to be a problem, we could host our own listserv. In addition, we could encourage more mail activity within peer groups (i.e., fifth grader to fifth grader, and mentor to mentor) by distributing e-mail address books.</td></tr>
<tr><td colspan="3">"Password crazies"</td></tr>
<tr><td>Observation/Problem: Despite the fact that we thought we made the dangers clear to the HS students, many of them chose passwords that they could not remember.</td><td>Insight: This was likely due primarily to the fact that they were presented with too many options in choosing their passwords: blank spaces and punctuation, etc. These students have little experience choosing and remembering PINs.</td><td>Possible Fix: We could ask them to pick a password and require that they enter the password on an actual keyboard (two times) before it is accepted. Alternatively, we could issue standard passwords to everyone that would be easy to remember or distribute printed cards.</td></tr>
<tr><td colspan="3">Lack of communication with mentors centrally and communication with students who had problems with e-mail</td></tr>
<tr><td>Observation/Problem: The high school students were never all in one class. Any time we needed to contact them as a group, we had only e-mail with which to reach them. Students who were having trouble with their e-mail could not be reached at all.</td><td>Insight: We need another method for contacting students. They have no regular access to phones, so we need more than a "hotline."</td><td>Possible Fix: A physical bulletin board centrally located at the high school where students could read posted messages and receive faxed personal messages.</td></tr>
</table>



Eudora has mailed your Ed story and has sent a copy of it to Elaine.

Fig 1. Three characters were introduced to provide the basis for students' mental models of the interface between the various software applications they were required to use.



Don't forget to:
save the file twice,

once in your folder on this machine (if you have one) and once on your floppy disk!!!

Fig 2. File management was identified as an issue and the source of much frustration to students. Dialogs were suggested to minimize the frequency with which files "disappeared."

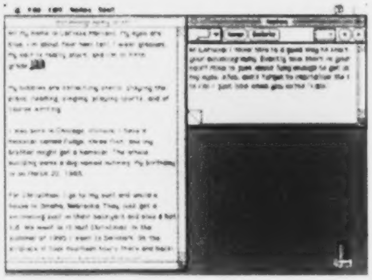
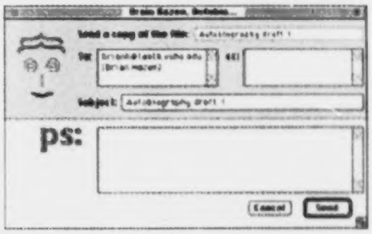


Fig 3. Next generation software specifications call for a separate window in which apprentices can read annotations embedded in their documents by their writing mentors.



Send a copy of the file: Autobiography draft 1

Subject: Autobiography draft 1

ps:

Fig 4. This mail window encourages students to distinguish between informal e-mail "notes" to their mentors/apprentices and the edited document that they are refining as a team.

BEST COPY AVAILABLE

Send comments or questions to:
jaym@id.iit.eduor eli@id.iit.edu

A Design Procedure for Training in Web

Ch. Metaxaki- Kossionides, Ass. Professor, Department of Informatics, University of Athens, Greece,
metaxaki@di.uoa.gr

Introduction

The use and integration of new information technologies in various environments adds value and increases the quality. The large amounts of linked information, the speed of access, and the friendliness of the interaction engage the learner. The Web used for information presentation and retrieval is strongly supporting the teacher and the student. The e-mail has changed our view of written communication.

The key issue investigated nowadays is the innovation of the learning environments created by these technologies.

The last three years we are engaged in local, national or international projects investigating relevant topics. We are more precisely investigating the pictorial communication, the interface design, the evaluation, the structure of content, the methodologies to design interactive presentations for training or awareness.

We have recently tried a design approach for familiarisation and training on e-mail [Metaxaki et al. 1996b]. It was quite well accepted by the trainees and gave good practical results [Kouroupetroglou et al. 1995]. In this paper we present a more systematic extension of that approach to the Web.

The main points of this approach are: the selection of a *mental model* the trainee already possesses, its cognitive analysis, the estimation of the differences and similarities between the *known* and the *new*. This approach creates cognitive skills and extends the possessed mental model [Mayer et al. 1992]. It is quite helpful even in the case of a faulty one. The mental model for the e-mail was the "communication" which was more or less an obvious selection. The selection of the mental model is very important, given that the design and the implementation are based on it. Concerning the Web this selection is neither obvious nor unique and is strongly dependent on the application environment. This makes the design procedure for the Web much more complicated, as our target is to present examples which can lead to the understanding of innovative use.

In the following we present the design procedure illustrated by an example.

The Design Considerations

The Web is commonly used for information presentation and *retrieval* or as a *dialogue forum*. The page design (images, texts, layout etc.) and the linking are important factors. In the case of the forum other factors must be taken into consideration i.e., the tracing of the dialogue, the interaction of the participants, the feeding of arguments, the conclusions' extraction and/or the tailoring of unsolved themes [Kobsa and Wahlster 1989], [Blattner and Dannenberg 1992]

For the e-mail, as we have already mentioned, the mental model was the communication. By the cognitive analysis of the relevant conceptual model we obtained the states, the connections and the dependencies. The differences and similarities between them and the parameters of the e-mail were presented. *A letter writing - sending by post* was the selected metaphor.

Concerning the Web, the selected mental model must first fit to the application environment. For the learning environment we can choose two mental models resembling the Web use: the *library* and the *dialogue*. Both of them are common and possessed by everyone. For simplicity reasons, we will continue with the dialogue model, as the same procedure is separate applied to both of them. The conceptual model of the selected mental model is cognitively analysed. We obtain a *set* of states, connections and dependencies, called *dialogue conceptual set*. The next step is the combination of this set with the Web, to find differences and similarities. We do not have a mental model for the Web, so we analyse its operational and formative characteristics. We obtain a *listing*, which can further split into branches. Each *branch* is formed by characteristics chosen to be appropriate for one use. We form the branch for the dialogue forum, called *WODIF*.

We enhance the dialogue conceptual set by the elements of the WODIF. This new set, containing the elements of both, the Web and the mental model, is the basis for the software design. What we need next, is to find the metaphor by which these elements will be shown clearly.
For evidence and better understanding we have selected *two metaphors*. The first is the metaphor of a dialogue between a *wise man* and his *disciples*. The second is a dialogue on a theme; i.e. an *historical event*. There are differences and similarities between them. In the first there is one *person of reference* for the questions and answers, the wise man. In the second a *given content* must be learned.
The screen layout, the dialogue evolution, the control flow, the interactions, the hypothesis testing are all scheduled and designed taken in consideration the elements of the enhanced set [Metaxaki 1994], [Metaxaki et al. 1996a], [Metaxaki et al. 1996c]. The product of this design procedure has so far given good practical results.

Discussion

In this paper we present a procedure to design an application for understanding and training on the Web. This procedure consists of a sequence of selections, estimations and analysis to obtain and combine the characteristics of both, the Web and the application.
The software designed by this procedure has quite a lot of advantages. The conceptual sets and the Web listings are supporting the designers. From the elements of the enhanced set the evaluation criteria can be found and tested. The user familiarizes on the same time with the certain topic and the Web innovation. We are currently working with metaphors of different content, to combine specific topics with Web use.

References

[Kobsa and Wahlster 1989] Kobsa, A., & Wahlster, W. (Eds) (1989). *User Models in Dialog Systems*. Berlin: Springer-Verlag.
[Blattner and Dannenberg 1992] Blattner, M., & Dannenberg, R. (Eds) (1992). *Multimedia Interface Design*. USA: ACM Press.
[Mayer et al. 1992]. Mayer, R. et al. (1992). The Instructive Animation: Helping Students Build Connections Between Words and Pictures in Multimedia Learning. *Journal of Educational Psychology*, 84 (4), 443-452.
[Metaxaki 1994] Metaxaki-Kossionides, Ch. (1994). Developing and Evaluating a Workstation for Teleteaching. *Telematics in Education: The European case*, (1994), (Ed) Wim Veen, Academic Centre Book, De Lier, 81-91.
[Kouroupetroglou et al. 1995] Kouroupetroglou, G. et al. (1995). A Generic Methodology and Instrument for Evaluating Interactive Multimedia. *Telematics for Education and Training*, (1994), 24-26 November, Duesseldorf, Germany, (1995), IOS Press, Amsterdam, 343-350.
[Metaxaki et al. 1996a] Metaxaki-Kossionides, Ch. et al. (1996). Investigation of Learning by Visual Information. *Technology and Communication: Catalyst for Educational Change*, (1996), 13th International Conference on Technology and Education, March 17-20, New Orleans, USA, 1, 344-346.
[Metaxaki et al. 1996b] Metaxaki-Kossionides, Ch. et al. (1996). A Multimedia based Educational Environment for Training Students in Network Technology. *Technology and Communication: Catalyst for Educational Change*, (1996), 13th International Conference on Technology and Education, March 17-20, New Orleans, USA, 1, 350-352.
[Metaxaki et al. 1996c] Metaxaki-Kossionides, Ch. et al. (1996). A Software Tool for Educational Research in Pictorial Communication. *ED-MEDA 96 Educational Multimedia and Hypermedia*, June 17-22, Boston-Mass., (1996), Association for the Advancement of Computing in Education, Charlottesville, 471-476.

Dynamic Educational Applications with Multimedia Databases: A Prototype

Sara Moshman
New Media Center
University of Nebraska-Lincoln
smoshman@unl.edu

The New Media Center at the University of Nebraska-Lincoln is a cutting-edge educational software development lab. One major work-in-progress is the creation of an Internet-accessible multimedia database server with client-side, java-based tools which allow the creation and storage of educational applications within the database that make dynamic use of the multimedia objects.

The database server is a dual-processor 200 MHz pentium pro, running Windows NT Server 4.0, Oracle 7, and Microsoft Internet web server. On the client side, a java applet interface, running under a java-enabled web browser, allows field searching of the objects. The results of a search are a series of thumbnails which match the query. Researchers can double-click on a thumbnail to show the object at actual size. The actual object files, currently only images, are stored externally to the database. Video and audio files will be soon addressed. Oracle Video Server and Progressive RealVideo products will being tested to stream video and audio content, as well as the java media APIs as soon as they are available. A jdbc/odbc bridge is currently being used for database access from java but research is being done with new jdbc drivers from Oracle. Networking infrastructure includes both ethernet and ATM, with the system being tested over both. Test results are not yet available.

The first educational application, written as java applets, provides the ability for researchers to create and store *conceptual paths* as they search the multimedia database. Conceptual paths are essentially fixed sequences of objects from the database which demonstrate or teach some concept. In other words, this allows the ability to create and store associations and interpretations about the data within the database. The researcher creates a conceptual path by copying the thumbnail of the desired object(s), and then pasting them into a conceptual path. These options are available from pull-down menus in the interface. Additional descriptive text can be added for each object in the conceptual path. The path can be saved with a title, the name of the creator, and a description and users can search them by those fields. Users can traverse any of the conceptual paths. When a user traverses a conceptual path, the multimedia object along with any descriptive text is displayed on the screen. Next and Previous actions are available.

As an example, a researcher might create a sequence of images from an architectural database on Frank Lloyd Wright to use for a lecture. This sequence can be stored in the database and the researcher can traverse the conceptual path and display the images as a presentation in class. Figure 1 shows sample screens for this example. At this level, the conceptual path is very much like a multimedia slide presentation - with the added value of retrieving the objects dynamically from the database at runtime. Copies of existing conceptual paths can be made and then edited to meet other needs. For example, another lecturer might copy the Wright sequence, add or delete objects, and store as a new conceptual path.

Security is provided by the database. Accounts are created for faculty with the privileges to create and edit conceptual paths. They can also add objects to the database and edit existing keywords and descriptions. They cannot delete objects from the database. Student accounts allow only searching and traversing of conceptual paths.

Issues of acceptable performance are being addressed in a number of ways. One is by using multi-threaded applets. While on the current presentation screen, another java applet is downloading the objects for the next one, so that when the user requests to advance, the information is instantly available. Database optimization will also be necessary, as will using compressed, low resolution images. Testing the application over an ATM network is also part of the performance testing.

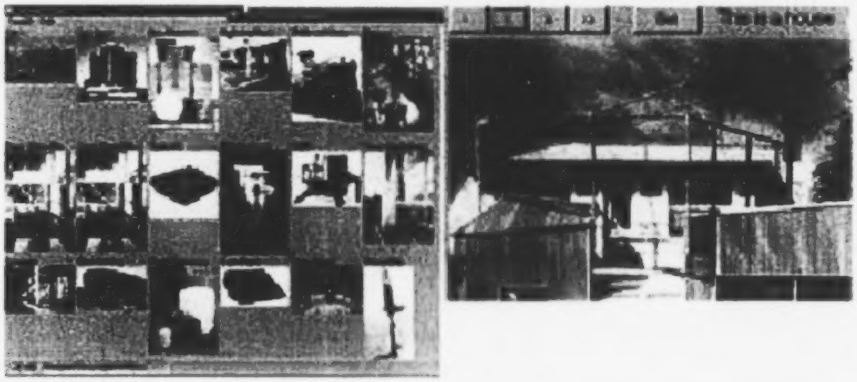


Figure 1. Search Screen Results

Presentation Screen for conceptual path

The second educational application, also based on conceptual paths and written in java applets, allows the creation of a multiple-choice question bank in the database. These questions also incorporate multimedia objects dynamically with the questions. Traversing this conceptual path will result in a multiple-choice quiz. Questions can be re-used in any conceptual path and dynamic random ordering of questions is possible.

The multimedia objects will become more complex. Currently, the objects are images, next video and audio, and eventually more complex objects such as an interactive simulation, will be incorporated. We are interested in object-oriented databases as our multimedia objects become more complex. The vision we are working toward is one of increasingly complex educational modules, or learning units, which can be dynamically created and customized to the learner at runtime.

BEST COPY AVAILABLE

OCS - An Online Conference Server

Michael Müldner, student, michael.muldner@acadiau.ca

Ivan Tomek, ivan.tomek@acadiau.ca

Jodrey School of Computer Science, Acadia University, Canada

Introduction

Organizing a conference involves a number of routine, time consuming, manual tasks. The following is a simplified description of the typical procedure of only a part of conference administration.

- Publish a call for papers with deadlines, a description of the theme and topics, etc..
- Finalize Program Committeee and reviewers, gather information on personal data.
- Authors send the required number of copies of paper submissions.
- Check submission and confirm its receipt, possibly notifying the author of problems.
- Upon deadline, organize submissions and distribute copies to matching reviewers.
- Send submissions and review forms to reviewers.
- Reviewers return reviews to conference organizers.
- Check and acknowledge reviews, possibly reporting problems.
- Upon deadline, collect and sort reviews, gather statistics, notify delinquent reviewers.
- Evaluate submitted papers, divide into accepted, rejected, and best papers.
- Send notifications to all authors with comments.
- Authors of accepted papers return edited submissions to conference organizers.
- Conference organizers review papers for adherence to guidelines etc.
- Organizers notify authors whose accepted papers have not yet been received.
- Upon deadline arrival, organizers organize received papers into proceedings.

It is obvious that the web can provide means for automating most of the tasks providing better response time, easy interaction, archiving, and monitoring. It can even provide new functions such as automated publication of submission and summaries on the conference site. Yet, a visit to the sites of current conferences (see a few selected sites in References for examples) and examination of the current procedures reveal that most conferences use the web only to publish calls for papers. We have not found a single conference that automates the process.

OCS is designed to do as much of the processing work on the server side as possible and minimize human intervention. Submissions are done by completing a World Wide Web form, selecting the conference paper filename, and submitting it by clicking a button. Reviewing is also done online. Reviewers can log onto OCS with a user name and password and review the papers that have been assigned to them, and leave their evaluations in a database for the administrator. The administrator has online control without having to enter the computing environment where OCS resides.

Design

The main conference page allows access to any modules in the system by recognizing which modules are present. It redirects the user through a web browser environment and provides navigation through the system. The paper submission page prompts the user for personal information, audio visual equipment, and a summary for the conference web site. A section of the form allows filling in the filename of the paper which is uploaded automatically to the server. Essential required information is checked and submission sent if complete.

Information extracted from the submission is assigned a submission number, a list of suggested reviewers created from keywords and areas of reviewer expertise, and added to the entry of the submission. The submission is added to submissions waiting for approval by the Administrator. The Administrator is informed of new submissions and each submission is shown with options to accept or reject. At the appropriate date, the reviewers are e-mailed an ID and password to allow access to the reviewer's page.

The Reviewer section of the conference site is accessible by ID and password. It displays the reviewer's list of unreviewed submissions and the reviewer may then read a submission and complete its review form. The information is added to the database and the administrator can access it and accept a review or notify the reviewer about problems. Upon deadline arrival, delinquent reviewers are notified and the administrator is reminded to make final review decisions. Review data is sorted according and when the committee makes decisions, authors are e-mailed the results.

Conclusion

The work on OCS has been undertaken as a course project and the software will be submitted for evaluation and used to develop a working software tool. The working software will be demonstrated at the conference.

With OCS, conference work can be greatly simplified, accelerated, and largely automated. Errors in processing are minimized and the process simplified. In the future, the tool will allow full handling of conference papers and processing both on the administrative and user side.

References

[CHI'97]	http://www.acm.org/sigchi/chi97/
[OOPSLA'97]	http://www.acm.org/sigplan/oopsla/
[TOOLS'97]	http://www.tools.com/
[AUSWEB '97]	http://ausweb.scu.edu.au/

The 15 Minute Series : Tools for Internet Trainers

Robin Murphy, Manager of Content & User Services
InterNIC Registration Services
Network Solutions, Inc.
Herndon, VA USA
robinm@internic.net

The amount of information and data stored, disseminated, and retrieved via the Internet grows by the second. As a result, the need to educate end-users in the methods and tools employed by the networked medium is also growing by the second. Some organizations are able to accommodate the budget and professional development challenges this growth presents, others find their resources strapped as they struggle to keep pace with expanding responsibilities. Rapid growth and frequent change in network technologies further increase the pressure on institutions and staff to stay one step ahead of their constituent audiences. This demand for education falls squarely on the intermediaries - libraries, computer services departments, faculty, information resource managers, and others - who play a role in providing information services to the end-users within their organizations.

In the fall of 1995, Network Solutions, Inc., the company that provides global registration services for the .com, .org, .net, .edu, and .gov top level domains, began an extended outreach program - aimed at the research and education community - under the auspices of its cooperative agreement with the National Science Foundation to provide Registration, Information, and Education Services for the Internet Network Information Center (InterNIC). The goal was to identify the most significant problems and challenges facing the research and education community in today's networked environment, and to determine appropriate ways in which the InterNIC, given its role and mission, might assist the research and education community with meeting those challenges. Several recurrent themes surfaced over the course of the outreach effort - most notably that organizations needed relief from the persistent yet critical task of training. In response, Network Solutions developed the 15 Minute Series.

The 15 Minute Series is a collection of Internet training materials provided as a service to the research and education community. The goal of the 15 Minute Series is to provide a resource that will assist this community in its efforts to incorporate and support the growing role of the Internet in day-to-day operations and activities. The project was publicly launched in September of 1996.

Developing a resource that would meet the needs of the "research and education" community was a tall order. While a single phrase, "research and education" is able to neatly capture the spirit of the community, it masks immense diversity in skill levels, technological infrastructure, and discipline specific interests. Clearly the 15 Minute Series would need to be general enough to work in a wide variety of training environments - dedicated Internet workstations, desktop delivery, presentations "for the road." In addition, the training materials would have to be flexible enough to allow a trainer to speak to the needs of a specific audience - audiences which might range from physics faculty to job hunting students and human resources staff. And, perhaps most difficult, the training materials would have to get the message across to absolute novices and seasoned Internet travelers alike, offering each information previously unknown.

How do you develop a resource that does all this for organizations that you will never set foot in? The approach that we took to the 15 Minute Series rested on three key principles: modularity, currency, and neutrality.

- Structuring the materials in a modular format provided flexibility. A trainer could select the modules that were appropriate for the training session, and use only those parts of those modules that were appropriate for the audience. Making a template available would enable trainers to extend the 15

Minute Series modules to speak to the specifics of the audience's environment; specifics that we - from our vantage point - could not possibly hope to cover in the materials.

- Regularly reviewing the training materials for currency allowed us to provide training materials that reflected current developments in networking technology and related issues, and ensured that the content of the training materials would remain accurate as time went by. Keeping the materials current also meant that the modules would be able to offer new information even on familiar topics.

- Finally, providing a variety of file formats, including platform independent formats such as HTML, allowed the modules to be used with a variety of technologies. Trainers using Macs or PCs, presentation software or the Web, network connections or non-networked machines would be able to use the materials.

The structure of the materials was one hurdle; presenting meaningful content in a concise and modular format presented yet another. To accommodate both objectives, the training modules were structured around a question and answer approach. Each training module would ask a specific Internet related question and provide an answer to the question in a consistent and succinct format. Graphics and analogies would be liberally used to help clarify difficult and complex technology concepts. The content would be thorough, yet remain general so as not to preclude its usefulness from one environment to the next. Each training module was designed to function on its own as a mini, self-contained presentation consisting of approximately 8-10 "slides" for use in formal training class situations. The simplicity of the language and thoroughness of the content, however, meant that the training modules were equally suited to individual, self paced training environments.

On August 31st, 1996, Network Solutions publicly released the 15 Minute Series via the Web and anonymous ftp. Each training module was available for downloading as a compressed Microsoft PowerPoint file and could be viewed via the Web. Debuting with 21 modules in its collection, the 15 Minute Series was an immediate critical success. Trainers could browse the collection in its entirety or by category, as well as search for modules on specific topics. Trainers wrote from all corners of the globe; the 15 Minute Series helped organizations with non-existent Internet training curricula, out of date materials, and even those with substantial collections of training aids.

In response to user feedback, we added a "packaged" HTML version of each module in October of 1996. We archived the 8 or 10 HTML files that represented each slide in each of the modules along with any graphic files, and then compressed the archive for efficient storage and speedy transmission. This step resulted in a third option - downloading the HTML version of the module - in addition to the two options already available (downloading as PowerPoint or viewing via the 15 Minute Series website). When properly decompressed and extracted on the trainer's end, a directory and file structure would automatically be created and all HTML and graphic files would be placed in the appropriate locations. The use of relative links in the HTML source code, when combined with the scripts used to create the compressed, archived HTML versions of the modules, enabled the modules and the source code to work on any machine. Trainers were only a click away from materials that were current, complete, and ready to use and distribute via their own web servers. Topics range from the basics such as electronic mail and the Web to cutting edge technologies and issues such as digital IDs and the next generation of the Internet Protocol. We are reaching trainers and audiences in higher education, foreign governments, and aircraft carriers in the Mediterranean Sea.

Taking the load from the shoulders of Internet trainers and demystifying the technology for end users are only two of the goals of the 15 Minute Series. The 15 Minute Series is dedicated to exploring and exploiting the potential the Web holds for increasing the number of people who are able to reap its benefits. By using the Web to distribute training materials that can be used with equal success in a networked or non-networked environment, the 15 Minute Series is able to reach a maximum number of current and future Internet users. By providing educational materials that increase general understanding of both the technical, historical, and societal underpinnings of the Internet, we hope to assist end users and trainers alike, and enable an even greater number of people to enjoy the benefits of this unprecedented medium.

WebView: A Multimedia Database Resource Integration and Search System over Web

Deepak Murthy and Aidong Zhang
Department of Computer Science
State University of New York at Buffalo
Buffalo, NY 14260

1 Introduction

Many applications involve accessing remote multimedia databases for the purpose of retrieving information which could be in the form of photographs, x-rays, scanned articles or satellite pictures. The access to such databases could be initiated by a client machine running a browser (such as Netscape or the Internet explorer) and allowing a typical user to issue image queries. These queries are processed by the browser and sent to the Web server which then selects the relevant target multimedia database site(s) and forwards the query to the database(s). The database involved searches for possible matches to the posed query and sends back the results to the Web server to be forwarded to the client. The selection of the most relevant database sites is based on the similarity of the query to the data in the sites.

This paper introduces a system termed WebView, as shown in Figure 1, which integrates multimedia databases for use in a Web-based environment. The three main components are multimedia database systems at Web sites, a meta-server consisting of a meta-search agent and a meta-database at the Web server, as well as a set of Web applications at Web clients. Using the Web as a medium to access multimedia sites involves accepting a user query in an acceptable form, selecting the appropriate sites, processing the query at the sites and presenting the results back to the user.

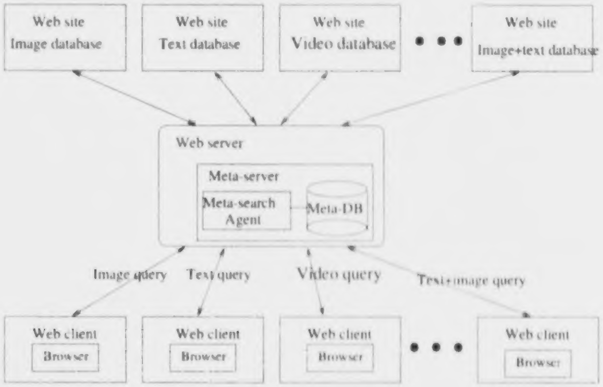

Web site
Image database
Web site
Text database
Web site
Video database
• • •
Web site
Image+text database
Web server
Meta-server
Meta-search Agent
Meta-DB
Image query
Text query
Video query
Text+image query
Web client
Browser
Web client
Browser
Web client
Browser
• • •
Web client
Browser

Figure 1: WebView: the Web-based integration of multimedia data resources.

2 System Features

A *standard 2-tier architecture* to construct a Java applet as a client for accessing a remote database server involves only the client and the remote database server. This requires extensive programming effort because a Java client must be implemented at the protocol level for vendor-specific databases. We design an *improved 3-tier architecture* which implements a stand-alone Java server (a Java application running in the Web server) as a gateway. The gateway passes the result and response messages between the applet and the

BEST COPY AVAILABLE

remote database server. This is more efficient in that the gateway server can be created as a stand-alone Java application with native client library functions wrapped in Java classes. It communicates as a server with the Java applet at one end while accessing the database server at the other end. We shall follow this approach since this is more flexible and portable.

The Web client is a typical user using a standard browser to invoke an HTML document which brings up a Java applet. The applet has the user interface implemented using the rich graphical features supported by the Java API and is used to obtain the user query. The user query is an image or a portion of an image and is accepted from the user interactively in that the user is allowed to display as well as select relevant portions of the image. The image query is then forwarded to the Web server which further routes it to the relevant database site(s). The final query results (names of the retrieved images) along with the names of the corresponding databases are displayed to the user and the image desired by the user (selected interactively) can be directly retrieved from the specific database using an associated CGI script.

The meta-database records information needed for database site selection. The meta-database is present in the Web server and organizes the information about remote database sites based on the type of queries they support and the types of media data they house. Given a set of databases at various sites, an initial meta-database is constructed from pre-defined templates (or icons) returned by the individual databases. Such templates contain meta-data about the data in the database, including the type of media data housed, expected query form, specialized algorithms supported and statistical data associated with each multimedia template. Data such as monetary cost and latency of database sites can also be stored to enable early pruning of costly sites. These templates can be periodically updated by the meta-search agent and relayed to component databases. The initial categorization of databases in the meta-database is used to direct queries to relevant sites. A record of recently returned responses and the associated queries is maintained at the server to avoid redundant searches.

Whenever a user query comes in, the meta-server must selectively forward it to the relevant databases (those having the highest potential to find images matching the query) for efficient searching. In order to achieve the above objective, templates consisting of sample icons are created to represent different classes of images corresponding to different databases. The information regarding these templates is stored in the meta-database. When a query comes in, the meta-server runs a local search to come up with matched templates. It then calculates the potential for each remote database by using the statistical information of the database and combining it with the similarity between the query and matched templates. This potential is used as a weight to rank the databases so that the top N potential databases can receive the query. The remote servers housing these databases then search for the query and return the matched images to the user. The functionality of the meta-server is split into two modules. The first module, *register module* is responsible for collecting and updating the information about the remote databases. The second module, *selection module* is responsible for making the decisions based on the information in the meta-database and is also responsible for routing the query to the remote databases. Both register and selection modules support dynamic configuration. Databases and templates can be added and modified dynamically and their new status indicated by the configuration files.

The environment is multi-threaded in order to support and follow up on queries from multiple users. Each client is handled from startup time to termination by a separate thread with the parent synchronizing between them. Java's socket API is used to provide connectivity between the applet, Web server, and remote servers.

3 Conclusion

We have presented a system for the integration of multimedia databases located at remote sites. This system includes the creation of multimedia databases, the meta-database, and the meta-search agent. A prototype to support Web-based multimedia information retrieval has been implemented using Java. The meta-database must be dynamically updated to prevent the redundancy of meta-data recorded due to frequent updates to databases. Refinement of an existing meta-database in response to component database updates is a difficult task. Since it is impractical to require the databases to report their status every time they are updated, the refinement of the meta-database can only be based on careful evaluation and validation of the query results. This has proven to be non-trivial, since we do not know whether the results are indeed relevant to the query without user input. We will pursue this research as part of the future work.

NetTest: An Integrated Web-Based Test Tool

Yun Ni Jianping Zhang Donald H. Cooley
Department of Computer Science
Utah State University
Logan, UT 84322-4205 USA
Jianping@zhang.cs.usu.edu

1. Introduction

As networking and multimedia technologies converge, the use of computer-based training and education has been experiencing a corresponding, dramatic increase. The success of computer-based instructional systems is partly attributable to their ability to evaluate the level of mastery the student has attained from instructions., i.e., the degree to which the student performance is congruent with the instructional objectives (Borich & Jemelka, 1981). Training and education should always be accompanied with assessment and testing. Unfortunately, administering, grading, and giving tests are labor intensive and time consuming tasks. Furthermore, traditional classroom tests are not flexible and students must attend a test at a fixed time and in a given place. For many types of tests, computers are an ideal tool to reduce labor and time requirements. Computer-based test technology has demonstrated its ability to carry out numerous tests efficiently and effectively. For these reasons, the use of computer-based testing technologies has recently experienced a significant increase recently.

We have implemented a computer-based test tool called UquTes (Universal Qualifying Test System) [Cooley & Zhang 1996] that delivers tests over a local area network. Problems arise when more than one test site is needed. One example of such multiple site tests is those for Civil Engineering Air Force personnel on which UquTes is currently being applied. These tests are being given at many different Air Force bases. In this case, each site must have its own test question databases and student database. Problems arise in maintaining the consistency of the test question databases and creating difficulties for collecting statistical data. Knowledge and technologies advance rapidly so that the test question databases need to be updated frequently. Also, a student can only take tests in one test site, because his/her record is only stored in one site.

There have been some efforts expended to develop Web-based testing systems. However, most of these Web-based test systems were developed as a part of a Web-based course [e.g., Bogley, et al. 1996] and include only multiple choice and fill-in-blank questions. Almost all of them are server-based, namely implemented using CGI scripts and HTML. This paper describes an integrated web-based test tool: NetTest. NetTest is Java-based and include six different types of questions. With NetTest and a Web browser, an instructor can create new test databases and enter new test questions; a student can take a test; and a test manager can perform various test management functions. By being Web-based, the system is not tired to a particular computer architecture.

2. NetTest

NetTest consists of four modules and two databases. The modules are the student module, the teacher module, the manager module, and the test generation module. The two databases are test the question database and the student, teacher, and manager database.

Using the student module, a student can:

take a test

view his/her information stored in the student database.

A test may be a locked test or a free test. To take a locked test at a test site, the student must ask the manager in that test site to unlock the test for him/her. A locked test is locked for all students until unlocked by a manager. The unlock signal is sent to the Web server and the test generation module randomly selects questions from the test database and all these selected questions are downloaded to the client machine. After the student finishes the test, the test is graded and the results (including the grade and answers to all questions) are sent back to the server. The record of the student in the student database and the statistical data for each question in the test question database are updated accordingly.

Using the teacher module, a teacher can:

enter questions into an existing test database using a Web browser;
browse test questions and select and edit questions to be included in a test using a Web browser;
grade tests;
access student information and test question statistics;
print hard copies of tests.

Using the manager module, a manager can:

create and delete accounts for students and teachers;
modify the information associated with a student or a teacher;
unlock a test for a student;
generate test statistics.

The test generation module runs on the Web server. It randomly selects test questions from a given test question database for a given student.

3. Test Questions

NetTest can accommodate the following six types of questions:

- Matching questions: Match items in column "A" with those shown in column "B".
- Sequencing questions: Place the shown steps in proper sequence.
- True/False questions
- Multiple choice questions (with one or more answers)
- Fill-in-the-blank questions
- Graphic questions
    1) Locate one or more given areas/points on a graphic.
    2) Recognize a given graphic object.
    3) Identify a number of graphic objects.
    4) Match each part of the items shown with the correct number on the graphic image.

4. Implementation

NetTest is fully implemented in Java. JDBC is used to access the databases. The database used is MSQL and the server is a HP workstation. The student, teacher, and manager modules run on a client machine inside a Web browser and the test generation module runs on the server. In comparison with the server-based system, the Java-based system has some advantages. First, the Java-based system reduces the load on the server because a significant part of the system runs on the client machine. This is important when a large number of users uses the system at the same time. Second, a better GUI can be implemented in the Java-based system and the GUI covers the whole screen so that the user does not feel that he/she is using a Web-based system. Third, the Java-based system is more interactive than the server-based system. Finally, some functions implemented in the Java-based system cannot be implemented in the server-based system. One example of such a function is the timer function that tells the student how much time is left for the test.

References

[Cooley & Zhang 1996] Cooley, D.H., & Zhang, J. "ConTest: A Multimedia Computer-Based Test Generation System," Proceedings of World Conference on Educational Multimedia and Hypermedia, June 1996.

[Dorich & Jemelka 1981] Dorich, G.D., & Jemelka, R.P., (1981) *Evaluation.* A chapter *in Computer-based instruction: a state-of-art assessment.* O.Neil, H. F. (Eds), Academic Press.

[Bogley et al. 1996] Bogley, A.B., Dorbolo, J., O.Robson, R., & Sechrest, J.A. "New pedagogies and tools for Web-based calculus," WebNet 96 - World Conference of the Web Society Proceedings, October 1996.

Acknowledgements

The work reported in this paper is supported in part by a grant from the National Science Foundation, CAD-9414210, and a grant from the Department of Education, FIPSE P116B40729. Thanks to Mr. Tzy-Tao Yang and Huan Zhao for implementing the manager and teacher modules.

The Quest for Major Forces Underpinning the Internet Worldwide Landscape

Enrico Nicolo'
Fondazione Ugo Bordoni, Research Division for Telecommunications Strategies, Roma, Italy, nic@fub.it

Bartolomeo Sapio
Fondazione Ugo Bordoni, Research Division for Telecommunications Strategies, Roma, Italy,
bsapio@fub.it

Introduction

The Internet is one of the most discussed topics in both technical and divulgative literature today. It is no easy task to browse publications *à-la-mode* or academic journals without stumbling over evocative words such as "virtual" or "cyberspace".

The Net phenomenon is a multifaceted, multilayered and multisectorial reality, yet most studies seem to focus only on particular or local aspects of this dynamic planetary scenario. Furthermore, very few studies cope analytically with the structures lying beneath the surface of this landscape. The need arises of interpreting and modeling this total-area panorama as a whole complex system, whose elements present mutual interrelations and evolve as a *web*, a global integrated entity.

This paper represents a step in this direction stemming from the identification of the most important factors of the planetary Internet scenario.

Factors Influencing the Evolution of the Internet

In a previous work we have identified 114 factors involved in the evolution of the Internet landscape, grouping them into four categories: technology, market, environment (context) and regulation [Nicolo' & Sapio 1997].

In order to effectively model this complex system, our first goal is now to reduce the number of variables, selecting the top ranking ones according to relevance criteria. We asked ten experts in the field of telecommunications to fill up a questionnaire, assessing for each factor an index of relevance in a predetermined range (from 1 to 5). The attention can be focused on the variables that present the highest average scores in each group [Tab. 1].

Technological factors
- Routers/bridges deployment
- Improvement of TCP/IP to support interactive multimedia
- Improvement of usability of multimedia systems
- Improvement of Internet search engines and indices
- Technological security
- Improvement of video compression/ decompression techniques
- Improvement of image compression/ decompression techniques
- Improvement of navigational tools
- ATM-based B-ISDN implementation
- Use of TCP/IP on ATM

Environmental factors
- National governments policies favouring information superhighways
- User attitudes towards multimedia systems
- User perception of the Internet
- User needs of multimedia services
- Political guarantee of access for all to telecommunications networks
- Educational and training needs
- Increase of consumer disposable income
- Human factors in computer mediated communications
- Strategic planning for enterprise information systems
- Diffusion of multimedia groupware

Market factors
- Number of Internet servers
- PCs diffusion
- Internet traffic
- Number of Internet users
- Number of Internetproviders
- Intranets diffusion
- Internet-oriented TV sets diffusion
- Presence of the games industry on the Internet
- Standards for browsers on the Internet
- Standards for video on the Internet

Regulatory factors
- Technical standardization activity
- Electronic crimes legislation
- Compliance to international standardization
- Privacy safeguard
- Definition of Intellectual Property Rights
- Internet advertising regulation
- Cross-border regulation
- Censorship regulation
- Antitrust regulation
- CATV regulation

Table 1: Top rankings for technological, market, environmental and regulatory variables

A particular care should also be given to variables showing a high degree of variance, meaning a strong disagreement among experts as regards their relevance. Here follow the ranking of the fifteen most relevant factors and the eight variables with the highest variances, after the evaluation of the experts [Tab. 2].

Ranking according to relevance means
- Number of Internet servers
- PCs diffusion
- Routers/bridges deployment
- Improvement of TCP/IP to support interactive multimedia
- Internet traffic
- Improvement of usability of multimedia systems
- Improvement of Internet search engines and indices
- Technological security
- Number of Internet users
- Improvement of video compression/ decompression techniques
- Number of Internet providers
- Intranets diffusion
- National governments policies favouring information superhighways
- User attitudes towards multimedia systems
- User perception of the Internet

Ranking according to relevance variances
- Network interworking for ATM-based B-ISDN
- ADSL diffusion
- Video on demand diffusion
- Intelligent network implementation
- Improvement of interfaces for disabled people
- Multimedia residential applications diffusion
- Human factors in computer mediated communications
- Cross-border regulation

Table 2: Top ranking variables on the basis of relevance

Future Research

This study is intended to serve as a step towards the analysis of the interactions between the factors identified. Given the set of significant factors influencing *global Internet scenarios*, it is possible to study their mutual impacts and rank them according to importance. Systemic and formal analyses can be carried out obtaining numerical results.

The chosen methodology is R-WISE (Reduced-Weighted Impact Structured Evaluation) [Sapio & Nicolo' 1997], a variant of WISE [Sapio & Antimi 1993], which is a quantitative method so far employed within specific sectorial analyses. R-WISE is a new method intended to reduce the complexity of data collection and selection during the process of scenario analysis.

It is worth noting that the selection of the top ranking variables in the Internet scenario presented in this paper is one of the steps towards the implementation of R-WISE in future research. This will provide strategic insight pointing out the structure of the relations among the variables in the Internet planetary landscape, describing it and reducing its complexity. It will also provide information about the capability of the different factors to influence the evolution of the considered scenario and to be influenced by it.

Conclusions

The paper has focused on the selection of the most relevant factors affecting the planetary Internet scenario out of a wider number of significant variables collected previously. Such a reduction activity, based on subjective evaluations by experts and accomplished through a quantitative ranking, is intended to be a first important step toward the analysis of the impacts that the considered factors exert upon each other in the global multimedia landscape dominated to a great extent by Web applications operated over the Net. Therefore, further research will be devoted to formalize and quantify such systemic interactions in order to provide information useful for decision makers in the fields of communications and information technology.

Acknowledgements

This work was carried out in the framework of the agreement between Telecom Italia and Fondazione Ugo Bordoni.

References

[Nicolo' & Sapio 1997] Nicolo', E., & Sapio, B. (1997). Factors shaping planetary multimedia scenarios: focus on the Internet in the era of global learning. *Orlando Multimedia '97*. Society for Applied Learning Technology, Kissimmee, FL.

[Sapio & Antimi 1993] Sapio, B. & Antimi, E. (1993). WISE (Weighted Impact Structured Evaluation): a new method for structural analysis, *ORSA/TIMS Joint National Meeting, 1993*, Operational Research Society of America and The Institute for Management Science, Phoenix, AZ.

[Sapio & Nicolo' 1997] Sapio B. (1997). Simplifying the experts' task during scenario analysis: the R-WISE (Reduced-Weighted Impact Structured Evaluation) method. *ISF '97 - The Seventeenth International Symposium on Forecasting*, International Institute of Forecasters, Barbados.

DISTRIBUTED CURRICULUM AUTHORING USING SGML

Roger Nkambou, Gilles Gauthier

Département d'informatique, Université du Québec à Montréal, Canada
Email: nkambou@safari.uqam.ca, gauthier.gilles@uqam.ca

INTRODUCTION

Instructional design is an important component of the teaching process [Merrill 93]. Instructional design requiring very much expertise and being a repetitive and tedious process, it would be helpful to empower the designer with tools that make his work easier. We are interested in the subject matter modeling part of the instructional design process. A conceptual framework for the subject matter has been proposed [Nkambou & Gauthier 96]. Also, an authoring environment (CREAM-Tool) consisting of tools (graphical editors, browser...) that enable the curriculum builder has been developed [Nkambou *et al.* 96]. A problem with this tool-based approach is that the instructional designer needs to be familiar with the interface. Also, this approach does not allow easily distance or distributed curriculum building and sharing. Our goal is to propose a new approach that consists of a language specification with its filters dedicated to curriculum building. This language will be as close as possible to the instructional designer vocabulary. The designer could then use any text editor to create a specification of his curriculum using the language. The role of the filters is to take the source file and generate a curriculum object from it.

CML SYNTAX SPECIFICATION

We use the *elm tree diagram* notation [Maler & El Andoloussi 96], to represent the DTD (Document Type Definition) of our Curriculum markup language (CML). From these diagrams, we build SGML code that implements the DTD. The DTD is the foundation of a SGML edifice. It can be used to control the authoring process and to provide information about a model to the software that formats them and processes them [Maler & El Andoloussi 96; Bradley 97]. For instance, HTML is a DTD that is used for delivery and presentation of documents over the WWW.

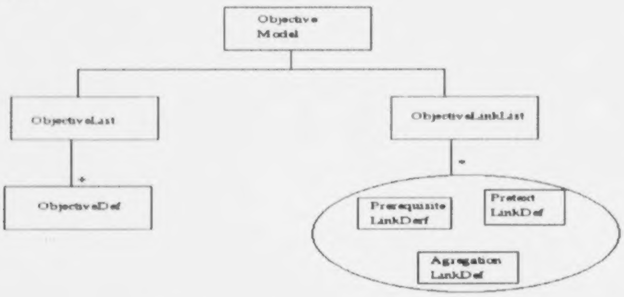


Figure 1: The objective model DTD

CREAM (Curriculum REpresentation and Acquisition Model) is the approach we used for curriculum representation in the context of an intelligent tutoring system (ITS) [Nkambou & Gauthier 96]. This approach models a subject matter from three points of view: the domain, pedagogical and didactic aspects. A curriculum according to CREAM approach is composed of a capability model, an instructional objective model, a resource model, a pedagogical model and a didactic model. The objective model DTD is represented by the figure 1. Figure 2 shows the SGML code corresponding to the objective model DTD.

```
<!ELEMENT ObjectiveModel (ObjectiveList, ObjectiveLinkList)>
<!ELEMENT ObjectiveList (ObjectiveDefList+)>
<!ELEMENT ObjectiveDef (SkillDescription, ObjectiveDomainDescription, Description?,
                        Context?, EvalRule?)>
<!ELEMENT ObjectiveLinkList (PrerequisioteLinkDef | AgregationLinkDef | PretextLinkDef)*>
```

Figure 2: SGML code implementing the objective model

BEST COPY AVAILABLE

THE OVERALL ARCHITECTURE

In the authoring system, two alternatives presents themselves (figure 3): authoring by using CREAM-tools (a curriculum-authoring environment in Smalltalk consisting of a toolkit to help with curriculum building) or authoring by using CML.

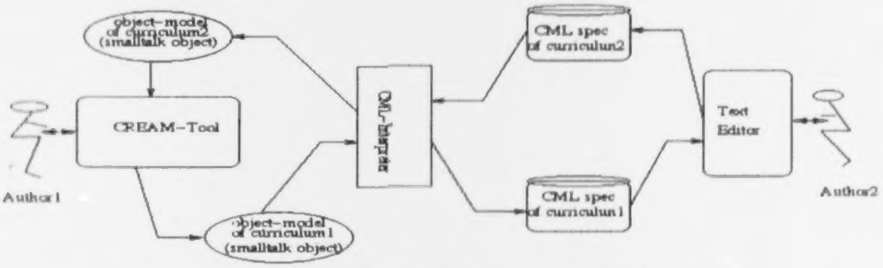


Figure 3: The authoring system functional architecture

Therefore, these two curriculum-authoring approaches are compatibles; for instance, let us consider two instructional designers developing a curriculum. The two authors (instructional designers) want to share their work. Author one, who possesses the CREAM-Tools software, uses that environment to build *Curriculum1*. Author two uses a text editor to create *Curriculum2* using CML.

CML filters are used in two ways:

- Take a CML document representing a curriculum, and generate a curriculum object model from the specification. The later could be a Java, C++ or Smalltalk object;
- Take a curriculum object model and produce the CML specification document of that object.

Thus, *Author2* will send his curriculum to Author1 by using *ftp* (file transfer protocol), or by other means. After receiving this, *Author1* will use the CML filters (CML-Interpreter (in figure 5)) to produce the Smalltalk object model of the received curriculum. This object model is then loaded in CREAM-Tool. He can now navigate in the obtained curriculum by using tools dedicated to that effect. Before send *curriculum1* to *author2*, *Author1* have to produce the CML specification of curriculum1 using CML filter, and then send the resulting specification by using *ftp*. When *author2* received it, he load it in his text editor and could bring all modification he wants.

We believe that the proposed authoring approach is a framework of distributed authoring since models in construction can be distributed among several authors simply by using ftp. Therefore, this distribution does not allow interactive discussion between authors involved in the courseware design process. This on-the-fly distributed authoring could be possible by integrating a co-operative model in the authoring process so as to enable authors to work in real time.

CONCLUSION

The specification of a generic and distributed authoring language for ITS curriculum building (CML) has been proposed and the CML DTD has been defined using the tree-diagram modeling approach and the corresponding SGML code have be derived. The authoring language obtained is a foundation to distributed authoring that will allow several author to collaborate in the curriculum building process. Our future work deals with the development of the CML filters and the experimentation of the approach by building real curriculums.

REFERENCES

[Merrill 93] Merrill, M.D. (1993). An integrated model for automating instructional design and delivery. In *Automating Instructional Design: Concept and Issues;* pp. 147-190, Englewood Cliffs, NJ: Educational Technology Publishers.

[Maler & El Andoloussi 96] Maler, E. & El Andoloussi, J (1996). *Developing SGML DTDs: From text to model markup.* NJ: Printice-Hall PRT, Inc.

[Bradley 97] Bradley, N. (1997). *The concise SGML companion.* New York: Addison-Wesley.

[Nkambou & Gauthier 96] Nkambou, R. & Gauthier, G. (1996). Un modèle de représentation du curriculum dans un STI. In *Lecture Notes in Computer Science 1086*, 420-429. Berlin: Springer-Verlag.

[Nkambou *et al.* 96] Nkambou, R., Gauthier, G. & Frasson, C. (1996). CREAM-Tools: an authoring environment for curriculum and course building in an ITS. In *Lecture Notes in Computer Science 1108*, 186-194, Berlin: Springer-Verlag.

BEST COPY AVAILABLE

Exploring Culture on the Web

Maritza M. Osuna, Educational Theory And Practice, University at Albany- SUNYA- New York, USA
mo0035@cnsvax.albany.edu

Carla Meskill, Educational Theory And Practice, University at Albany- SUNYA- New York, USA
meskill@cnsvax.albany.edu

Over the last twenty-seven years, the issue of teaching culture as part of foreign language (FL) learning has triggered an outburst of scholarly work. This rapid growth in interest paralleled attention given to the communicative nature of language and the attainment of communicative competence [Savignon, 1972; Canale and Swain, 1980]. Increased understandings about culture at a theoretical level have not however resulted in increased practical implementation. As a result, many foreign language learners demonstrate very limited cultural knowledge [Sadow, 1987]. Several explanations for this situation have been proffered: FL instructors' lack of current cultural knowledge, inappropriate training [Nostrand, 1989], and deficient textbooks that frequently become the guiding force behind FL syllabi and curricula. Textbooks tend to bestow one-sided views of the target culture through piece-meal approaches [Kramsh, 1988; Lafayette 1988].

The quest to develop sociolinguistic competence in its proper cultural context has brought the use of authentic materials to the FL classroom. This pilot study examined uses of Internet resources as authentic foreign language materials for teaching and learning culture.

Materials and Method

This pilot employed six Spanish activities using the Internet. These activities, which integrate Spanish culture and language, were tightly coordinated with the class textbook ÀHabla espa-ol? M ndez-Faith et al.,1993]. The following abbreviated activity illustrates the type of task created using Softguide Madrid (http://www.softdoc.es).

Planning a family trip to Madrid

Your family is planning a vacation to Madrid, Spain. Since you speak Spanish, you are in charge of finding as much information as possible about the city. (1) Your family will need accommodations. Find names and fees for different kinds of hotels (i.e., luxury, moderate, hostels, etc.), then decide where you will stay. If you were traveling with friends, where would you stay? (2) You will need to eat. However, your father wants authentic Spanish food, your mother is vegetarian, you enjoy eating light food, and your little brothers want fast food. Find restaurants to please everyone in your family. (3) Your family will like to take some day trips around Madrid. Find a place to visit, find the train number you will need to take, and decide if this is a good place for your little brothers. (4) Your family wants to visit the Prado Museum. Find the location and hours of operation. You probably want to visit when there is free admission. When is it?, etc.

A post-activity assessment questionnaire concerning attitudes and perceived learning outcomes accompanied each activity.

Subjects

Subjects were thirteen undergraduate students enrolled in the first trimester of Elementary Spanish. Except for one, all subjects were computer literate with prior Internet experience.

Results and Discussion

Results demonstrate that the Internet is an excellent tool for teaching foreign language and culture. Data showed that 88% of the subjects reported increased knowledge of Spanish language and culture. This finding is remarkable because all activities reported high marks for increased knowledge, with higher gains as activities progressed. Explanations for these results may be found in the type of task, appropriate balance of language and culture, and authenticity provided by the Internet. Tasks emphasized realistic language use rather than language rules. Hence, task completion required a more hands-on approach with tacit knowledge of linguistic forms. Combining the text's cultural information with the language of the chapter proved to be an ideal departure point to broaden the students learning experience. Furthermore, students were engaged in the activities because they saw them as an integral part of the class, not as addenda. As in real life, language and culture remained together, without overpowering each other to foster authentic communication.

Results for culture learning separated from language learning were also very good. Eighty-one percent of the subjects reported that cultural learning was occurring. Subjects were able to focus on culture (separated from language) because language forms were presented implicitly and did not demand added attention from the users. When subjects were asked about learning language separated from culture, 77% of them reported gains. Interestingly, some subjects did not explicitly recognize that language learning was taking place. But this finding is not surprising because of what is known about implicit-explicit language teaching and learning [Shaffer, 1989; Scott, 1990; Green and Hecht, 1992; DeKeyser, 1995].

Even though separation of language and culture still produced satisfactory results, the best outcome was obtained when language and culture were integrated (i.e., 88%). This finding strengthens the importance of teaching language and culture in context, a point that cannot be overemphasized. Data presented here demonstrate that if FL students are to become successful learners, integration of language and culture is pivotal.

Technology outcomes are very promising. When asked about attitude towards the medium, 85% of the subjects reported having a positive attitude, in spite of occasional technical difficulties found when completing activities. The high mark for satisfaction (85%) with the medium (i.e., after each activity) was increased to one 100% in a retrospective survey. This is a very exciting result, especially for FL teachers searching instructional activities that will increase time on task. Data collected showed that there is a direct correlation between satisfaction and level of interest. If, as subjects reported, using the Web makes the class more interesting, they will be willing to spend more time performing a task or browsing over other information connected to it.

Subject interest was aroused by what the medium has to offer: current, interesting, varied, and useful information backed by multi modal attributes that proffer text, sound, and visuals [Meskill, 1996]. One of these characteristics, to which subjects categorically referred and enjoyed, was the visual text. Exposure to visual text proved to be a great asset in increasing positive attitudes and cultural learning. According to Monroe [Monroe 1993], education has overlooked the relevance of visual stimuli over verbal and analytical skills. Hence most graphic inclusion in instructional design is founded in instincts, not in principles [Rakes, 1996]. Nevertheless, it has been demonstrated that visuals can be employed to aid learning and foster positive attitudes [Poohkay and Szabo, 1995], and visual stimuli can also become memory-assisting devices [Stickels and Schwartz, 1987]. Data presented here strongly suggests that inclusion of visual text for FL instruction is highly desirable. Nevertheless, this integration needs to be thoughtfully planned to produce positive outcomes, such as accelerating learning, increasing learning efficiency, and facilitating retention [Rakes, 1996].

Conclusion

This pilot study integrating Spanish language and culture using the Internet affirms that the medium is a valuable tool for foreign language and culture learning. Technology seems to be especially beneficial in promoting cultural learning. Subjects also reported numerous advantages of the Internet over other media and instructional tools. There is much to be discovered in the application of new technology to the FL classroom. The results of this pilot implementation should encourage other FL instructors to become active participants of that breathtaking world that is just a screen away.

References

Alvarez Evans, G. & Gonz‡lez, O.G. (1993). Reading "inside" the lines: An adventure in developing cultural understanding. Foreign Language Annals, 26 (1), 39-48.

Ballman, T. (1996). Integrating vocabulary, grammar and culture: A model five-day communicative lesson plan. Foreign Language Annals, 29 (6), 37-44.

[Canale and Swain 1980] Canale, M. & Swain M. (1980). Theoretical basis of communicative approaches to second language teaching and testing. Applied Linguistics, 1,1-47.

Garc'a, C. (1989). Development of sociolinguistic skills in Spanish: Sample activities. Foreign Language Annals, 22 (4), 357-366.

[DeKeyser, 1995] DeKeyser, R.(1995). Learning second language rules: An experiment with a miniature linguistic system. Studies in Second Language Acquisition, 17, 379-403.

Garc'a, C. (1991). Using authentic reading texts to discover underlying sociocultural information. Foreign Language Annals, 24 (6), 515-526.

[Green and Hecht, 1992] Green, P. S. and Hecht, K. (1992). Implicit and explicit grammar: An empirical study. Applied Linguistics, 13 (2), 169-184.

[Kramsh, 1988] Kramsh, C. J. (1988). The cultural discourse of foreign language textbooks. In A. J. Singerman (Ed.), Towards a new integration of language and culture (pp. 63-88). Northeast Conference on the Teaching of Foreign Language. Middlebury, VT: Northeast Conference.

[Lafayette, 1988] Lafayette, R. C. (1988). Integrating the teaching of culture into the foreign language classroom. In A. J. Singerman (Ed.), Towards a new integration of language and culture (pp. 47-62). Northeast Conference on the Teaching of Foreign Language. Middlebury, VT: Northeast Conference.

[M ndez-Faith et al. ,1993] M ndez -Faith, T., Mcvey Gill, M. & Mayne Kienzle, B. (1993). (5th ed.) ÀHabla espa–ol? Orlando, FLA: Harcourt Brace.

[Meskill, 1996] Meskill, C. (1996). Listening skills development through multimedia. Journal of educational multimedia and hypermedia, 5 (2), 179-201.

[Monroe, 1993] Monroe, S. (1993). Doors and windows of my mind: Thresholds of visual thinking. (ERIC Document Reproduction Service No. ED 360 213).

[Nostrand, 1989] Nostrand, H. L. (1989). The beginning teacher's cultural competence: Goal and strategy. Foreign Language Annals, 22 (2), 189-193.

Pennington, M.C. (in press). The computer and the non-native writer: A natural partnership. Cresskill, NJ: Hampton Press.

[Poohkay and Szabo, 1995] Poohkay, B. and Szabo M. (1995). Effects of animation and visuals on learning high school mathematics. (ERIC Document Reproduction Service No. ED 380 122).

[Rakes, 1996] Rakes, G. C. (1996). Visuals in instructional design. Performance and instruction, 35 (3), 30-32.

[Sadow, 1987] Sadow, A. S. (1987). Experiential techniques that promote cross-cultural awareness. Foreign Language Annals, 20 (1), 25-30.

[Savignon, 1972] Savignon, S. J. (1972). Communicative competence: An experiment in foreign language teaching. Philadelphia, PA: Center for Curriculum Development.

[Scott, 1990] Scott, V. M. (1990). Explicit and implicit grammar teaching strategies: New empirical data. French Review, 63 (5), 779-789.

[Shaffer, 1989] Shaffer, C. (1989). A comparison of inductive and deductive approaches to teaching foreign languages. Modern Language Journal, 73 (5), 395-402.

[Stickels and Schwartz, 1987] Stickels, L. and Schwartz, M. (1987). Memory hooks: clues for language retention. (ERIC Document Reproduction Service No. ED 337 013)

Web Access to CD-ROM Databases:
A Model for Easy, Equitable Access to Government Information

Carolyn Ottow
The Valley Library
Oregon State University
United States of America
ottowc@ccmail.orst.edu

The Government Information Sharing Project (http://govinfo.kerr.orst.edu), developed at Oregon State University Library, is an interactive web site providing access to several useful statistical databases including the 1990 Census, School District Data Book, USA Counties, the Census of Agriculture, and the Economic Census. The Project began in 1993 with funding from the U.S. Department of Education. The original intent was to provide access to government information on CD-ROM to remote users in Oregon and the Northwest. With the emergence of the WWW and freely accessible development tools, the site has become a valuable resource for people across the U.S. and around the world, serving over 6000 unique hosts each week. Many of the databases available on the Government Information Sharing Project (GISP) web site are electronic counterparts to standard reference sources that have been available in academic and larger public libraries for years. Providing Internet access to them has demonstrated that the WWW can help to make access to government information more equitable by removing or reducing geographic, technical and intellectual barriers.

Because government information generally is not copyrighted, it has long been a major source of the informational content of the Internet. In the early 1990s, census information extracted from libraries' CD-ROMs was accessible at a few gopher sites, but often only the most basic summary reports or reports for a given geographic area were available. Using an interface that interacts directly with the CD, the GISP site can generate reports from a much greater amount of information. The data on the CDs is stored in dBase format. By adapting DButil software developed at Lawrence Berkeley Laboratory [Merrill 1996], programmers on the project wrote software to read and format the data files. CGI scripts were written for the web pages' interactive forms. When a user queries the site, the programs extract the data directly from the CDs which are in drives attached to the web server. One exception is the 1990 Census data. Because of the large number of CDs (over 60), the dBase files were extracted, subsetted, and stored on the server's hard drive.

At the time we began to develop the site, web servers at the Lawrence Berkeley Laboratory (http://parep2.lbl.gov/cdrom/lookup) [Merrill et al. 1995] and the University of Virginia Social Sciences Data Center (http://www.lib.virginia.edu/socsci/) [Bergen 1995] were using similar kinds of interactive interfaces to government CDs, but they were geared toward researchers and academic users. The goal of the GISP is to make the data more easily accessible for the general public, and to provide a means of outreach for the library. The site is designed to be as user friendly as possible, and the use of technical jargon is consciously avoided. Users do not need to know the specific "tape file" that the data came from or have any knowledge of statistics. Information from all databases is easily selected using a consistent interface of interactive forms, scrollable lists and clickable maps. Keyword searching is available within each database, allowing users to pinpoint statistics on specific topics. The complete documentation from each CD is available in HTML to explain how the data was compiled, define terms, and identify sources and authority of the data. Original help screens were also written to provide context sensitive help in navigating and querying the data sets.

Another important aspect of the site is its compliance with standards to reduce barriers to the information. Testing for multi-browser compatibility, including text only browsers, assures access to the widest audience possible. Browsers that are used to query the site are identified by usage tacking software mounted on the server. Avoidance of proprietary software formats and use of only standard HTML also helps to assure broad access.

Despite trends among agencies to use the sale of electronic government information to increase revenue, the experience of the GISP clearly shows that equitable access to federal information can be achieved efficiently and at low cost. Information compiled and published at tax-payer expense should be freely available. Charging usage fees would mean that the information is accessible only by those who can afford to pay and would thus restrict access. By making it more accessible, new applications for the information will be found. The freely available demographic, economic and educational statistics on the GISP have been used by journalists, teachers, farmers, community planners and senior citizens, some of whom had not previously been aware of these resources. Another indication of the new application of the information is the recognition the site has received from various web rating systems, including K-12 educational sites such as the Eisenhower Clearinghouse for Mathematics and Science Education and Learning in Motion. Given this success in disseminating government statistics it is hoped that the GISP can be seen as a model for providing easy, broad and equitable access to government information.

References

[Bergen 995] Bergen, P. (1995). Interactive Access to Geographic Information Systems and Numeric Data on the World Wide Web. *Journal of Academic Librarianship*, 21 (4), 303-308.

[Merrill 1996] Merrill, D. (1996 January 14). *DBUTIL: Unix Software for Accessing 1990 Census Data on CD-ROM.* Berkeley, CA: Author. [Posted on the World Wide Web]
http://cedr.lbl.gov/cdrom/doc/dbutil.html

[Merrill et al. 1995] Merrill, D., Parker N., Gey F., and Stuber C. (1995). The University of California CD-ROM Information System. *Communications of the ACM*, 38 (4), 51-52.

Acknowledgments

The Government Information Sharing Project was originally funded by the Department of Education, College Library Technology and Cooperation Grants Program. Continuing support is provided by Information Services Department, Oregon State University. The web site is the result of the hard work and talents of the entire GISP team. I would especially like to thank ron stillinger, Jacquelyn Miller, Kathy Howell, Qing Yang, and Stephen Mosely for their technical expertise and dedication to the project.

The Intranet Support Problem

John R. Potter
John Deere Waterloo Works, USA
PotterJohnR@waterloo.deere.com

Introduction

The great strength of today's intranets - its wide variety of communication services - is also its greatest weakness. Since the services are provided using a wide variety of hardware platforms, operating systems, protocols, and applications that must all interact; it is increasingly difficult to insure high availability. Nonetheless, I.S. staffs use their intranet to provide services that are critical to running today's enterprise. (If you wonder if intranets really provide *critical* services, think what would happen in your enterprise if its email service was unavailable for a day.) This paper arose from an organization that uses a large number of diverse services, from the mundane Domain Name Service (DNS) to elaborate, web-based, on-line transaction processing applications. An issue that we face, as do most groups that support highly developed intranets, is this: how does an organization assure high availability of their intranet services? In this paper we present fundamental assumptions in intranetworking, demonstrate how they lead to the "Intranet Support Problem" and propose a form that a solution may take.

Assumptions

We feel there are some assumptions that are fair to make in regards to most highly developed intranets. They are:

1) Users of intranets insist on an ever larger number of diverse and elaborate services, e.g., printing, file, email, scheduling, web serving, OLTP, etc.;
2) Since builders of intranets believe in "open systems", they usually have no qualms about purchasing solutions in the form of software and hardware from different vendors;
3) It is the nature of tcp/ip that high-level protocols rely on low-level protocols. Hence it is axiomatic that intranet application services interact with lower level services, and it is fair to assume that these interactions have data and time complexity/constraints;
4) Intranet builders are always tinkering with services, i.e., adding, reconfiguring, or retiring services.
5) Resolving an intranet service problem is often difficult. Typically finding the problem cause and a suitable solution requires a person who is highly knowledgeable of all layers of the tcp/ip protocol stack as well as the configuration of the services in the particular intranet.
6) Downtime for intranet services is a critical problem for the enterprise, and quick solutions are required. This often results in the "lets reboot and see what happens" solution. A more thorough understanding of the problem and solution requires the ability to dig through mountains of details and log files. Since this solution often takes a considerable amount of time it is difficult to convince management and staff alike that it is worthwhile.

Problem Statement

The intranet support problem is this:

> *Intranet services often do not behave in ways that we expect because the complexity of the software lends itself to subtle bugs and because the services have increasingly complex interactions with other services. Failure to resolve this uncertain behavior will most likely lead to insupportable intranets in the near future.*

The intranet support problem is faced every time we tinker with a service. We have to answer obvious questions like: Will this version of operating system work with this new service? And less obvious questions like: I have the same platform, the same version of operating system, the same version of web browser, but some user's browsers consistently crash after viewing this image, and others never do, why is that?

Frustratingly, the intranet support problem is also faced during periods where no changes have been made to the intranet services. A failure may be the result of an operating system failure on a machine seemingly unrelated to the service in question, or other times a change in the usage (number of users, frequency of messages, etc.) may cause unexpected behavior.

Solution Form

Before proposing a solution we note a better solution that is not available to us because of an assumption listed above. In general we feel that lightweight services with that meet basic needs would be easier to support, i.e., have less complicated interactions with other services, and be less likely to contain bugs. These lightweight services are eliminated by our assumption that the user community needs elaborate (i.e., heavyweight) services.

We propose not a solution, but a form the solution may take. It has two parts:

1) Test suite - A collection of hardware and/or software tools that isolate the service from its environment (i.e., other intranet services) and exercise its inputs and verifies its outputs.

2) Monitor suite - A collection of hardware and/or software tools that monitor a service while it is in operation in the intranet. The suite watches inputs (checking that they are of the correct number and type), throughput (checking for acceptable flow), and outputs (verifying consistency with profiles of known good behavior). Since there would be many monitor suites for an intranet, it would be useful if they reported back to a single station.

The test suite would be used before releasing a service to the enterprise. It would help the service installers have confidence that the service is configured correctly and that it can indeed provide the level of service expected. Additionally, it could also be used at times of major system maintenance to verify that the service is still in "tune".

The monitor suite would continually watch the service, verifying that it is available and providing expected levels of service. During a failure the monitor suite would indicate which lower layer service was not responding correctly.

Any solution for the intranet problem must resolve this paradox: the monitor suite is, in effect, an intranet service. How do we know that our monitor suite is really running correctly? To prevent other intranet service failures from causing the monitor to fail it may require separate hardware for monitoring and a separate network for communicating among the various monitor platforms. This is not unlike using a second, separate Ethernet for fire and security applications, rather than sharing the traditional data network.

Conclusion

In this short paper we have listed the assumptions made regarding highly developed intranets which lead to the intranet support problem. We defined the intranet support problem and predicted that increasing complexity of interactions may lead to insupportable intranets, i.e., intranets that can not be made to be highly available. A suggestion for a form of a solution was described. A complete monitor solution would be welcome, though a simple solution that provides even limited functionality that is highly available would be more acceptable than a highly complex system that promises to monitor everything.

Logan: Formal Telemeeting Document Support for the World-Wide Web

Gitesh K. Raikundalia
Human-Computer Communications Centre
Bond University, Gold Coast
Queensland, Australia, 4229.
gitesh@dstc.Bond.edu.au

Introduction

World-Wide Web (Web) tools for document management during formal, synchronous, distributed electronic meetings are lacking at present. This paper presents such a tool, *Logan* [Raikundalia & Rees 95] [Rees & Raikundalia 96], that supports highly effective creation, presentation, access and navigation of the formal documents of agenda and minutes. The tool's user interfaces have been developed iteratively through numerous experimental meetings. It supports telemeetings, consisting of pre-meeting, meeting and post-meeting phases. Discussion is structured during the pre-meeting phase via an agenda that motivates meeting participants to discuss items in order to achieve some purpose. These meetings are executed in a sequence, forming meeting chains. Discussion of meeting phases can be found in various places such as [Bergmann & Mudge 94] and sequences of meetings in [Morrison 93].

Logan provides user interfaces for a *secretarius* (a participant assigned administrative duties of the meeting) and participants to develop the agenda. It handles agenda creation and arranges appropriate linking to documents. It provides Web mechanisms for secretarius development of minutes post-meeting in an item-by-item fashion. Another main contribution Logan makes is to perform analyses of meeting transcripts of discussion (log analysis) to form derived documents, called *derivatives* (these are discussed in the abovementioned citations).

Agenda

The agenda is the driving force of the meeting. A Logan agenda is modelled on traditional meeting agendas thereby making the transition to telemeetings simpler. In addition to elements found in traditional agendas, such as date, meeting purpose, participants, agenda items, the Logan agenda:

- is informative of instructions for entering the meeting, such as suggested times for tool login
- is presentable and well-structured as a Web page (exploiting HTML elements strongly, such as image maps, tables, varying fonts and lists)
- provides convenient accessibility to Web pages and documents, such as immediate access to the minutes of the last meeting and access to the participant's agenda contributions page (discussed later).

An agenda is developed from *contributions* by participants of the meeting (during the pre-meeting phase). A contribution is a set of details suggested by a participant regarding a single agenda item. The mechanism used is that of *secretarius moderation* whereby the secretarius will determine which contributions are relevant and appropriate. Those selected by the secretarius as "accepted" are automatically added by Logan to the agenda. Those rejected are added by Logan to a *rejected participant agenda contributions page* (RACP) for that meeting. The mechanism works such:

1. The secretarius requests contributions from participants by an email generated and sent via Logan.
2. Participants voluntarily contribute to the agenda. To do this, they view the current states of both the agenda and the rejected contributions (which are anonymous) on the *participant's agenda contributions page* (PACP) to guide them in contribution formulation. They can view unacceptable types of contributions with associated reasons provided by the secretarius.
3. The secretarius views all contributions, either accepting a contribution (Logan adds it to the agenda) or rejecting it (Logan adds it to the RACP). The mechanism then goes to step 2 when further contributions are made by participants. On rejection of a contribution, the participant may reformulate and resubmit it.

These steps occur until a final date for contribution submission (indicated in an email sent to all participants) when the agenda is finalised.

Minutes

The minutes record for each agenda item the outcomes, decisions and actions on participants succeeding from the meeting. An item may have one or more associated decisions. For each decision, an action is able to be recorded, therefore producing an action list. Each action consists of the participant to carry out the action, the action itself and a due date by when the action must be completed. Experimentation revealed that a decision and its associated action needed to be kept physically close together. Again, tables proved reliable in structuring the information as needed.

Automation of minutes creation is achieved using an interface consisting of the derivative *verbatim minutes* and a form for which details for one item are submitted. Logan allows the secretarius to fill in details about the outcomes, decisions and actions by using information from the verbatim minutes. Alternatively, s/he can easily select options indicating that the item was not covered, and if necessary to carry over the item to the next meeting. When all necessary details are supplied the minute is submitted, and a fresh form is loaded preparing the secretarius to supply details for the next item. In this way, all items are covered and minutes are created item-by-item as found to be the best manner through experimentation.

Findings

Two series of experiments, *exp1* and *exp2*, were carried out. From the 16 meetings, some of the findings are:

1. The necessity to partition a document into a title frame at the top and the actual content as the remainder of the page. Losing one's way in the tool while navigating during a meeting of intensive discussion is highly probable without this separation. Hence, the context of the page in use must always be maintained.
2. The necessity to compose a page of more than one document or interface, but with three components at the most (excluding the title frame). More than three components would be less manageable—scrolling the components becomes excessive, text is harder to follow and HTML elements like tables become distorted.
3. The necessity of maximally using horizontal space. Some pages can potentially become very long due to the need to present a large amount of information or provide form elements. It was found that a participant was more likely to miss information or elements nearer the bottom of the page due to lengthiness of a page.
4. The presentability of information due to HTML table elements. Information becomes ordered and clear, which is much-needed for participant satisfaction during a meeting. Tables also assist in achieving point 3 since text or form elements may be aligned horizontally as cells in the same row.

References

[Bergmann and Mudge, 94] Bergmann, N.W., & Mudge, J.C. (1994). Automated assistance for the telemeeting lifecycle. *ACM Conference on Computer-supported Cooperative Work, Oct. 1994*, Association for Computing Machinery, North Carolina, USA. 373-384.

[Morrison 93] Morrison, J. (1993). Team memory: information management for business teams, *Twenty-Sixth Hawaii International Conference on System Sciences, 1993,* Wailea, Hawaii, 4, 122-131.

[Raikundalia and Rees, 95] Raikundalia, G.K., & Rees, M.J. (1995). Exploiting the World-Wide Web for electronic meeting document analysis and management. *AUUG '95 & Asia Pacific World-Wide Web '95 Conference, Sept. 1995,* Sydney, Australia. 82-94.

[Rees and Raikundalia, 96] Rees, M.J., & Raikundalia, G.K. (1996). Deriving user interface functionality for document support during electronic meetings. *First Asia-Pacific Conference on Computer-Human Interaction, June 1996,* Singapore. 92-104.

Acknowledgements

The work described in this paper was undertaken with equipment supplied by the Cooperative Research Centre for Distributed Systems Technology (DSTC), Australia.

InterWeb : A Web-Based System for Collaborative Learning and Work

Riccardo Rizzo, Santi Scimeca
Istituto di Tecnologie Didattiche e Formative, Italian National Research Council
Via Ugo La Malfa, 153 - 90146 PALERMO - ITALIA
rizzo@itdfmac.itdf.pa.cnr.it, scimeca@itdfnext.itdf.pa.cnr.it

Introduction

The newest metodologies for supporting learning and working processes are more and more involved with the use of telematics resources. Distance education system of third generation *(on-line education)* [Harasim 1989], are mostly based on Computer Conferencing (CC) technology, by which a tight interaction between participants (tutors, experts and learners) and a structured, off-line communication process are ensured [Jonassen. et al. 1993]. There is involved a process of collaborative learning [Kaye 1994], stimulated by two kind of interaction:

- learner-content.
- learner-learner

In the first case, the interaction is supported by the help of experts and it is facilitated by the CC system, which structures the content of a course in *stages* and *modules* gerachically and cronologically significant. The second kind of interaction is important for developing a collaborative learning process, and the role of tutors is to facilitate the communication between the learners in the conferences. When we pass from collaborative learning to collaborative work we notice that CC systems lack of functionality. They are not so practical when people from different places have to project and create a negotiated object, like a written document or a structured text. Moreover, when a virtual community has to build up a common document using a CC system is often obliged to switch to other desktop applications. The object of discussion is never visible on a shared ground, so there is a need for a tool which facilitates an *integrated communication* and the *manipulation* of the object of work. For *integrated communication* we intend a communication process which involves text, images and hypertextuality; for *manipulation* we intend the possibilty to collaboratevely discuss and update the document without switching to other applications.

We can see the World Wide Web as a place where it is possible to find, in addition to great information resources, shared informations. Web-conferencing (WebC) systems are gaining credibility for becoming the communication tools of the next generation information technology. In this systems topics can be structured in conferences and threads, in which messages are organised [Pampili 1996]. But only one aspect of the Web seems to be used: messages are seldom composed by multimedia elements. InterWeb has been thought to use the hypertextual flexibility, the versatility and the *information openness* of the WWW. By means of InterWeb:

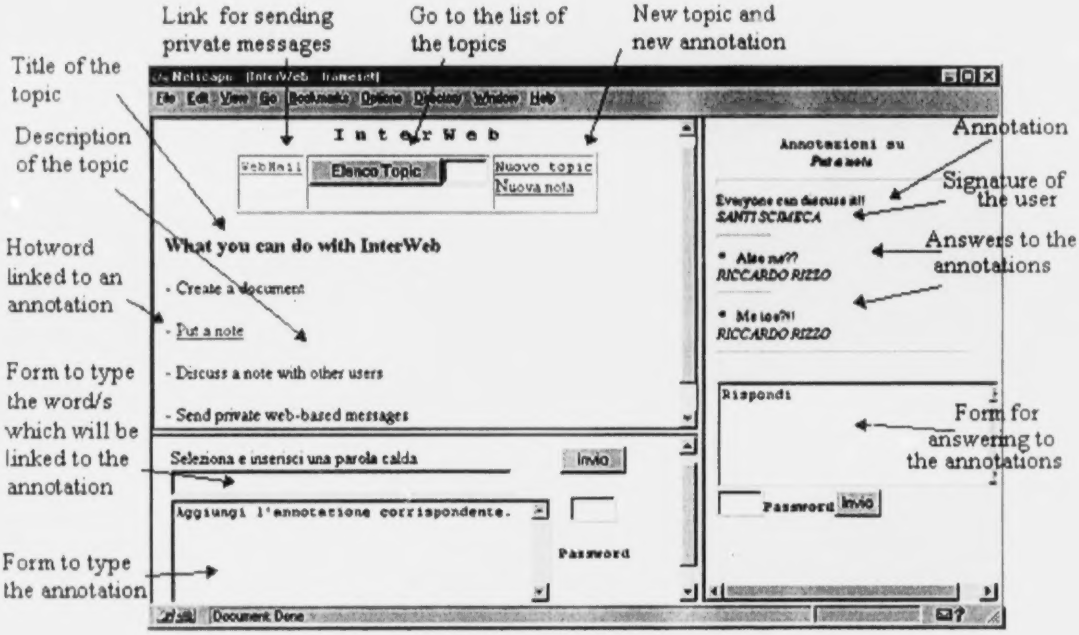
- you can put on the net a document which contains all the elements of the World Wide Web; this document is on a web server, so it is shared by all the partecipiants of the communication process;
- every user can link document's words with annotations;
- every user can answer these notes, creating a sort of web-conference;
- you can link your document to other web pages.

Technically speaking, InterWeb is made up by several CGIs written in Perl and has been implemented on a PC Pentium100 Mhz with 16Mb RAM.

How It Works

InterWeb uses the solution of the *frames* (see the latest releases of Netscape and Explorer). When you connect to a site which incorporates InterWeb, you can access the system by typing name and password. If you are a registered user the answer of the system is a page with a list of all the open discussions which signals also the new topics and annotations. You can also log in as a guest, but you can not write.

From this page a user can either access each discussion or go to a page where it is possible to insert a new topic. If you choose to start a discussion about a new topic, you can type (or paste) it in a specific form. When you send the form or, from the previous page, join a discussion, you access a page divided into three frames [Fig. 1]. In the biggest portion of the screen there is the shared document which can contain the hotwords linked with the annotations. If hotwords are present, clicking on each one there appears the frame on the right with the referring annotations, plus a form where you can insert an answer. The frame at the bottom of the page contains the form by which you can associate a word of the document to an annotation.



What you can do with InterWeb

- Create a document
- Put a note
- Discuss a note with other users
- Send private web-based messages

Seleziona e inserisci una parola calda

Aggiungi l'annotazione corrispondente.

Annotazioni su Put a note

Everyone can discuss all!
SANTI SCIMECA

- Also me??
RICCARDO RIZZO

- Me too?!
RICCARDO RIZZO

Rispondi

Link for sending private messages
Go to the list of the topics
New topic and new annotation
Title of the topic
Description of the topic
Hotword linked to an annotation
Form to type the word/s which will be linked to the annotation
Form to type the annotation
Annotation
Signature of the user
Answers to the annotations
Form for answering to the annotations
Netscape - [InterWeb - frameset]
File Edit View Go Bookmarks Options Directory Window Help
InterWeb
WebMail
Elenco Topic
Nuovo topic
Nuova nota
Invio
Password
Invio
Document Done

Figure 1: The main page of InterWeb

InterWeb rapresents a special kind of web-conferencing system, more oriented to collaborative work processes. From the standpoint of the information structure of the environment, InterWeb is much more open than a CC environment. You can develop connections from every element of the document by creating hypertextual links, and discuss such annotations. In this way every user can negotiate modifications on the document, which can be updated in every moment. Every participant has the same version of the document visualized on the screen, and can discuss about each element with precise, *clickable* references. Moreover, InterWeb gains advantages from the multimedia features of the World Wide Web, with the possibility to visualize text associated with images. A tutor can start a discussion about every topic of the course and answer to student questions which are linked, as annotations, to the main document.

We are planning to use the system as a communication support for distance education and *on-line education* courses, especially in those subjects where the richness of the information structure is a predominant aspect.

References

[Harasim 1989] Harasim L.M., (1989). Online Education: a New Domain. *Mindweave: Communication, Computers and Distance Education*, (Mason, R.D. e Kaye A.R. eds) , Oxford, Pergamon Press.

[Jonassen. et al. 1993].Jonassen, D. et al., (1993). "A manifesto for a constructivist approach to uses of technology in higher education" in *Designing Environments for Constructive Learning*. (Duffy, T.M. et al. eds), NATO ASI series, F105, Heidelberg, Springer-Verlag, 231-247.

BEST COPY AVAILABLE

[Kaye 1994] Kaye A., (1994). Apprendimento collaborativo basato su computer. *TD - Rivista di Tecnologie Didattiche*, n.4, Edizioni Menabò, Tortona, Italia, 9-20.

[Pampili 1996] Pampili T., (1996). Groupware. *PCMagazine*, October, http://www.pcmag.com /issues/1517/pcmg0027.htm.

Electronic Transcripts in the Age of Standards

Robert Roberts
Oregon State System of Higher Education
PASS Project
robert_roberts@ccmail.uoregon.edu

Robby Robson
Oregon State University
Department of Mathematics and University Honors College
robby@math.orst.edu

Introduction

In 1991 the Oregon legislature called for the elimination of grades and "seat-time" as a means of measuring progress through school. Two series of standardized State benchmarks lead students to a "Certificate of Initial Mastery" (CIM) in 10th grade and a set of standards for the awarding of a further "Certificate of Advanced Mastery" (CAM). In 1993 the Oregon State System of Higher Education (OSSHE) responded by adopting a new system for admissions to all public institutions of higher learning, to be phased in by the turn of the century. The change to proficiency-based admissions in higher education poses the problems of re-thinking the notion of a high school transcript and building a model for the collection, storage, transmittal, and retrieval of standards-based educational data.

Proficiency-based Admission Standards System (PASS)

The Proficiency-based Admissions Standards System (PASS) mandates that students be admitted to OSSHE universities on the basis of demonstrating that they have met proficiency standards in all of a set of 44 proficiencies in 6 content and 9 process areas. Student proficiency is verified by three primary methods: 1) Standardized multiple choice tests, including SAT II tests and tests created for the CIM and CAM; 2) Common performance assessments, which are tasks or projects with state-wide definitions and common scoring criteria; and 3) Teachers applying common scoring criteria to student work samples. A document crucial to admissions and subsequent academic advising decisions in this system is the electronic transcript.

The Electronic Transcript

The goal of the electronic transcript is to create a path from the verification of proficiencies to admissions decisions. The PASS project has formed an " Electronic Transcript Design Specification Team" and in doing so has sought and received the input of admissions officers, representatives from all levels of public education, from private education, from industry, and from persons responsible for data processing at the school, district, county, and state levels. Three of the main issues we have identified and addressed are: Levels of Data, Transcript versus Application Materials, and User Interfaces.

Levels of Data

The single statistic currently used for college admission is the Grade Point Average (GPA). The high school transcript reduces the student's record to a list of grades in classes in a single-layered document. A web-based electronic transcript in contrast can be multi-layered.

In the standards-based system, we identify three main levels of data. The top level consists of summative scores or binary decisions concerning whether or not a student has met a given standard or collection of standards. In the case of PASS, the scores assigned to each of 44 proficiencies are at this level. The second level consists of the verifications which went into making the decisions. In practice we are interested in knowing only the most basic information at this level: who, what, when, and how was the verification done. At the third level is the student work which was assessed. Data and entries can be hyperlinks or portals to more detailed information concerning students achievement of proficiency. We see an opportunity for electronic transcripts to be linked to multi-purpose student electronic portfolios. Not only could these portfolios represent the verified work of student proficiency, but they could also be used for non-college bound students seeking employment. In between the second and third level there could also be a set of completed "scoring guides" which were used to make the proficiency verification.

Transcript versus Application Materials

It is helpful to focus on the difference between a transcript and an application. Traditionally, the high school transcript reflects only grades achieved in classes. In a standards-based system, the transcript records the student's success in achieving the standards. By extension, this includes PASS proficiencies. Components which are not generated by a high school or which are provided as part of an application to a single university, may be part of the admission's package but are not part of a transcript (e.g., SAT and AP scores, CIM and CAM scores, relevant work or internship information, and admission's essays). OSSHE has at least provisionally made the progressive decision to keep all standardized applications data for on-demand accessibility by OSSHE institutions. The model is to create electronic paths into an OSSHE database. One set of paths emanates at school districts and counties and carries PASS and other data. Other sets will emanate from the Educational Testing Service and various external sources (school, district, or private digital storehouses).

User Web-Based Interfaces

Our vision includes three web interfaces: 1) Teachers or others entering verification data: 2) Admissions officers, advisors, and families who are "end users"; and 3) Applicants. The reasons for this include universality (platform independence and user familiarity), ease of use, and flexibility (including the ability to upgrade).

The current need for supplemental university material addressing student competency can be eliminated by having a reliable proficiency-based admissions system. If universities believe that external evaluations are reliable and are based on a sufficient body of evidence, then there is little need to review but a small subset of that evidence for the purposes of admissions. The student will also be able to upload essays (addressing, for example, goals or interest in a particular institution) and provide other electronically available supplemental information via a web interface (e.g., testing agency test score results stored in a central OSSHE database).

Using an authenticated and secured web interface, the student will be able to submit a single application to OSSHE. This application will authorize OSSHE to request data from the relevant school, school district, or county and also authorize OSSHE to release the data to those OSSHE institutions named by the applicant. The ability to do this is conferred by existing software used by OSSHE institutions (e.g., Banner which accepts data in EDI format via SPEEDE Express). Banner has a web interface which we are already using for on-line admissions.

Described in this short paper are some of the challenges in designing and implementing electronic transcripts used primarily for admission to higher education institutions. The solutions we are prototyping are robust, scalable, and inexpensive. The PASS Electronic Transcript Design Specification Team is at: http://iq.orst.edu/pass/tds/. It is one small but very important and innovative piece in the age of standards-based education.

Productivity Tools for Creating Web-based Presentations

William Rosener, Ph.D.
Northeastern State University, USA
rosener@cherokee.nsuok.edu

Julie Little-McGill, Ed.D.
University of Tennessee, USA
jklittle@utk.edu

John Rose
University of Tennessee, USA
rose@cs.utk.edu

Utilizing the Web in the educational process does not require radical restructuring of current practice. Whether instruction is offered in the contiguous or non-contiguous learning environment, all instructors engage in a standard set of activities. Some of these activities include delivering information, issuing assignments, managing grades, generating tests, offering resources, and communicating with students. The tools that support these activities, the delivery of material and administration of information, are called productivity tools.

Generally, these tools exist as stand alone software applications - presentation slide shows, syllabus generators, grade managers, test generators, resource cd-roms, and communications packages. For Web-based course delivery, there are productivity tools that act much in the same. A few examples include an assignment collector [HREF 1], a syllabus generator [HREF 2], a grade averager [HREF 3], a photo album [HREF 4], learning questionnaires [HREF 5], and online dictionaries [HREF 6]. There are also several online conferencing tools including NetForum [HREF 7] and E-Pub [HREF 8]. Therefore utilizing Web-based productivity tools is merely the continuation of current practice in another medium.

However tools still need to be developed that allow new presentations to be created, especially for those instructors who cannot afford to purchase commercial presentation software. For these instructors, we are developing a system called *WebPresenter* [HREF 9] which uses an interactive CGI script to to develop professional looking presentations viewable from the Web. Using a basic browser with no plug-ins attached, users can quickly create the pages for an HTML presentation. Unlike other HTML editors, this system is being designed and developed exclusively for developing Web-based presentations. Everything from the backgrounds to the clip art has been chosen for their presentational value. Before discussing the *WebPresenter* system, we will first examine the benefits of Web-based presentations. We recommend visiting the following URL if you have never attended or viewed a Web-based presentation before (Web Design Tips [HREF 10]).

Benefits of Web-based Presentations

Rationale: The Web continues to explode - in size and access. As a result, it will likely become the dominant computing technology in the delivery of resources for contiguous and non-contiguous learning environments. Whether it is used as a sole method of delivery or to augment a traditional face-to-face course, instructors must develop systematic ways to organize and present the information available on this vast network. To insure instructors will use the Web at a paralleled growth rate, Web-based tools are being developed to help them organize information and conduct classroom activities.

Right now, there is no standard model for delivering educational opportunities over the Web. Instructors can select from a variety of productivity tools to create a customized model - from presentation to communication to management uses. However, an examination of these tools can be overwhelming for instructors interested in, but lacking experience, utilizing the Web for instruction. The most difficult aspect of integrating the Web in the teaching practice is to determine **where** it fits and **how** to use it. Identifying points-of-entry will enable instructors to introduce themselves, and in many cases their students, to Web-based educational experiences.

To find an entry point, instructors should examine what they are currently doing in their traditional classrooms to find activities that would easily translate into Web-based delivery. One activity currently used in the contiguous classroom, and an easy point of entry, is the presentation. Presenting information with slides, transparencies or using the board has been central to the traditional learning experience. Beyond it being a convenience, there are other advantages to using Web-based presentations.

Hyperlink to Other Resources

The Web's outstanding feature is the incredible amount of information available to users. Couple that with instructor expertise in particular subject areas and one finds a built-in capability to organize vast resources for student manipulation. The advantage to using Web-based presentations is the ability to hyperlink to other Web resources. The instructor who scours the Web looking for collateral resources to reinforce the concepts presented, offers their students enhanced learning opportunities.

Integrate Multimedia

Typically, instructors utilize a variety of media in their classrooms - text, graphics, audio, and video. Expanding on the aforementioned, web-based presentations can integrate such by hyperlinking to web-based multimedia resources. While most presentation software packages allow for the integration of various media, storage requirements can inhibit portability and/or display capabilities. Web-based presentations can link concepts to web sites with multimedia files, further enhancing the medium's visualization and automation capabilities while easing storage requirements. In addition, a plethora of multimedia resources available on the Web could expand beyond budget-limited in-house resources.

Easy Distribution

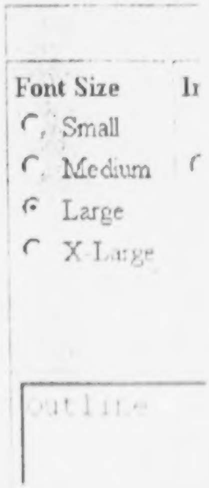
Distribution of information is central to the educational process - contiguous or non-contiguous. Typically, instructors use presentations to introduce new material. However, their use can be extended to review material or as self-paced instruction for individual learners. Using web-based presentations to review material or for self-paced instruction can ease time and place constraints. If students are able to experience new material or review the same outside of the time and place constraints of the traditional classroom, they are able to extend their learning opportunities. Obviously removing the constraints of time and place is paramount for the non-contiguous learner.

WebPresenter

WebPresenter is an on-line interactive CGI script which allows user to develop, with ease and speed, professional looking presentations for the Web. Using a basic browser with no plug-ins attached, users can quickly create the pages for an HTML presentation. Unlike other HTML editors, our system is being developed exclusively for developing Web-based presentations. The clip art, backgrounds, and font colors have all been selected and designed for their presentation value. WebPresenter can be viewed and tested at the following URL:

http://www.it.utk.edu/itc/tools/presentations/

To create a presentation the user first specifies a background by either selecting one of the default options or providing a user defined background. Next the user begins entering textural information. This information can be entered as individual items or points. For each item, selection and pull-down menus can be used to indicate the font size, amount of indentation, font color, and bullet type as shown in Figure 1.



Font Size
○ Small
○ Medium
● Large
○ X-Large

Figure 1

When the user is finished entering information, they can press the "Display Slide" button and an document similar to Figure 2 is displayed in their browser. Using the browser's "Save As" option, this document can then be saved and upload to a webserver.

BEST COPY AVAILABLE

Netscap
File
Edit
Back
Forw
Netsc
What's N
Doc

Figure 2

Beyond Individual Productivity Tools

As mentioned earlier, Web-based productivity tools enable instructors to translate current

BEST COPY AVAILABLE

educational activities for Web delivery. New and intermediate users are encouraged to utilize these tools for partial integration of the Web into their teaching practice. This in turn will stimulate thinking for restructuring courses for complete Web delivery.

References

HREF 1 - *Assignment Collector* (http://wwwtools.cityu.edu.hk/collect.htm).

HREF 2 - *Syllabus Generator* (http://www.it.utk.edu/itc/tools/syllabus/).

HREF 3 - *The Letter Grade Averager* (http://www.mindspring.com/~waus1/javastuff/gradeavg.htm).

HREF 4 - *Photo Album* (http://wwwtools.cityu.edu.hk/album.htm).

HREF 5 - *Learning Questionnaire* (http://www.it.utk.edu/itc/tools/questionnaires/).

HREF 6 - *Online Dictionaries* (http://www.bucknell.edu/~rbeard/diction.html).

HREF 7 - *NetForum* (http://www.biostat.wisc.edu/netforum/Docs/createforum.html).

HREF 8 - *E-Pub* (http://www.e-pub.com/epub/home.html).

HREF 9 - *WebPresenter* (http://www.it.utk.edu/itc/tools/presentations/).

HREF 10 - *Web Design Tips* (http://www.it.utk.edu/itc/papers/webnet/tips/).

Nääsnetti - a 3D Media Village for Social and Informational Uses of Media

Mr. Timo Saari, Director of Research,
R&D Communications, Aamulehti Group Ltd., Finland
email: timo.saari@aamulehti.mailnet.fi

Social uses of media, collaboration and telepresence are hot topics in the World Wide Web nowdays. This is realised in the emergence of numerous 3D-worlds with chat functions. A 3D- world is fascinating because the producer of the application actually produces a platform for creating content. With this kind of research focus we created within the Näasnetti multimedia research -project, the Media Village, a 3D- environment build by the R&D of Aamulehti Group Ltd., a major mass communications company in Finland.

The Media Village is not an electronical newspaper but rather an evolving interface with access to various services. It consists of a 3D- VRML-1-based space with different buildings and rooms with graphical icons on the walls representing a set of services as links. The social use of networked multimedia is realised by providing a text based chat service and an Internet Phone application. Each user in the 3D- Media Village is represented by an avatar and can talk to others IRC-style.

The uniqueness of Näasnetti Media Village at the time in the end of 1996 was that it combined the well established 3D-chat cultures and applications with an ordinary interface for information retrieval. The result was a communicative and social 3D-space which could also be used for accessing services. The hypothesis for creating such an environment was to see whether any kind of social use of content (e.g. chatting about the links or information on the walls) would emerge in addition to the "normal" chat.

The Media Village Application

The 3D- space in the Media Village works in a Netscape 3.0 or newer WWW-browser with Live 3D-plug-in for VRML. The Village consists of buildings: a media house, a school, a communal services building, an entertainment center and a shopping center. The buildings act as a classification of services and include rooms with services as links on the walls or objects such as radios or tv´s. The services included electronical newspapers, a news wire service, an Internet radio station, video-on-demand-services, multimedial journalistic articles, games, advertisements, interactive comics and textual chat services. We also added an Internet Phone application to the set of services. Mouse or keyboard are used to navigate the 3D-space. During the test phases in Tampere, the Media Village was accessed in a Pentium pc-computer with a broadband Internet connection, such as LAN.

The main content of the Media Village was the electronical newspapers of Aamulehti Group Ltd.: Iltalehti, our evening daily (www.iltalehti.fi), Aamulehti, our flagship regional daily (www.aamulehti.fi) and Kauppalehti, our business daily (www.kauppalehti.fi). We also offered a personalised, real time news service provided by our news agency, Short Message Services. Video-on-demand services were based on the offerings of the local cable-tv-station Tampereen Tietoverkko. We also used the audio services of our local and Internet radio station Radio Moro (www.alexpress.fi/moro). Experimental contents included for example journalistic articles with video added to common web pages as a value-adding component to the whole of the text and picture -based story. In them we experimented for instance with interactive interviews with ready made questions and edited answers on video and three-level structural journalistical hierarchies for an easy reading experience. Interactive comics were based on Macromedia Shockwave technology with audio and animation. We produced two full weeks of a comic story based on everyday life of an imaginary family in Tampere. The plot was linear for five days a week, Monday through Friday with five episodes published daily while the stroryline and dramaturgy was developed. On Friday the users could choose inbetween two different endings for the story. The selection was carried out by a voting program built on our server and operated through a web page. Each week of the comic story was manuscripted as a separate entity. In the shopping center we created 3D commercials for a housebuilding company. For chat we used a localised version of Cyberhub by Black Sun Interactive and the server was located in Münich, Germany. For video-on-demand services we used both MPEG-1 and AVI-format and a mediaserver with an ATM connection in Tampere Telephone Company. For audioservices we utilized the server pool of Telecom Finland and their Medianet-service with the software of Xing Technologies. Also a common WWW-server was used.

The Usability Study and The Reception Study of the Media Village

The usability study was conducted by the Usability Lab of the Department of Computer Science in the University of Tampere. The results of the usability study show that there still are major usability problems with VRML-technology. Users ran into problems such as slow updating of the screen, application crashes and lacking properties of the 3D-browser. In navigation the main aspect is to be able to understand the world well enough to be able to make use of the space metaphor as an interface. In addition to 3D-browsing users preferred also different kinds of shortcuts to be able to access different parts of the 3D-world easier than by browsing. These kinds of aims for navigation are widely used in ordinary hypermedia applications and the need for them seems to be obvious also in VRML-worlds.

When a 3D-world is used as an interface metaphor the key question is what kind of a metaphor it is. Users were confused about the functionalities of the world. This means that instead of just building a cool 3D-world one should carefully consider what excactly is the focus of the 3D-visualisation. In the case of the world acting as an interface 3D is used to represent and visualise the structure of the application. If this is done with a real life metaphor the users will expect to be able to open doors, look out the window or so. One solution for not confusing the users could be to use a more abstract metaphor, for example a futuristic city that does not give the user so many chances to mix the browsing experience with one's experience of the physical world. Visibility and scope of one's "virtual vision" into the world should also be wide enough so that users would not have to spin around the world too frequently.

The field trial of the Media Village was carried out in December 1996. The users were in local schools, universities, student apartments and Aamulehti Group. All trial sites had a broadband connection to Internet, mainly LAN with an ATM backbone. 357 users registered for the experiment and a survey about their preferences and experiences was put out during the second week of the trial on a web page. 77 users answered the 46-question survey and the age division was fairly even ranging from 11 years to 53 years of age. The co-planning of the survey and analysis of the survey data was carried out by the Center for the Journalism Research and Development in the Department of Journalism and Mass Communication in the University of Tampere.

The reception study produced interesting results of which a few are mentioned here. Basically the users were more interested in the chat than in the interface properties of the application. Most users described their browsing style in the world as "wandering around" which seems to validate that 3D-worlds could actually be more than just interfaces, rather they are experiences. The youngest users preferred hanging around-type browsing whereas the older users preferred information search and the use of services. Not surprisingly the youngest users liked entertainment services and older ones information services. Electronical newspapers and especially local information were considered of great importance. All in all the users had a positive attitude towards the world despite the common technical problems. The social use of the information services provided was not widely detected, rather it seemed that chat and hanging around was a social event and the use of information services was more a personal act.

In 1997 the Media Village part of the Nääsnetti project has continued on the basis of the results of 1996 and we have been focusing our research on Java supported VRML-2-worlds with shared objects, user's object manipulation capabilities, shared multimedia information and shared broadcast mediastreams.

A Web-Based Training and Advising System for Software Libraries

Ahmed Seffah(1) and Cao Lieu Nguyen(2)
(1) Computer Research Institute of Montreal
1801, Avenue McGill College Suite 800, Montreal (Quebec) Canada H3A 2N4
Phone: (514) 840-1235 extension 5437 Fax: (514) 398-1234
Email: aseffah@crim.ca Web: http://www.crim.ca
(2) Computer Science Department
Université du Québec à Montréal
CP 8888, Montreal, Quebec (Canada)
Phone: (514) 987-3000 extension: 3227 Fax: (514) 987-4501
Email: nguyen.cao_lieu@uqam.ca Web: http://www.uqam.ca

Introduction

Nowadays, many software libraries, generally called APIs (Application Programming Interface), are available to increase productivity and improve software quality. However, in most case theses libraries are not effectively and efficiently used or/and reused. [Desmarais 93] shows that only 50% (±20%) of the services of most software applications is ever mastered. The other half is either not useful to the specific needs of each user, or he or she has never had the time or made the effort of mastering it. Evidence from our day-to-day experience suggests that the latter is also true for software libraries.

Our position is that the following are the main reasons of this problem:

- Learning how to use well a software library is a long and very hard task, even for an experienced software programmer. In order to meet a particular user's needs, services must combined; several combinations are possible. The effectiveness of each structure depends directly on the user's level of skills and his or her experience about using the library.
- [Carroll 96] explains that the difficulties of learning object-oriented design and software are compounded by the fact that expert programmers often have high confidence in their ability to learn new programming techniques, and find long and steep learning curve frustrating and demoralizing. . Evidence from our day-to-day experience suggests that the latter is also true for software libraries. As a consequence, there is a substantial drop in productivity when programmers start to use a new API.
- Software libraries offer a large range of powerful but complex services. However, most of them are only useful in specific domain, under some conditions and hence cover only a part of a user's needs. Besides, this problem becomes more complicated, because a programmer often uses more than one library.
- Generally developers don't like to consult documents (user's guide, reference manual) and often refuse to do so. Furthermore in many online help systems, the help messages are often incomplete and not well adapted to the user's context.

In this paper, we present a new training and advising system that tries to offer a coherent solution to the mentioned problems.

An Architecture based on Internet Tools

The proposed system is based on the typical Internet clients/server infrastructure and the intelligent help system architecture [Kearsley 88]. The main features of the system are:

- Beside a tool for browsing through information about services, the system includes tools for advising and training.
- A unique object-oriented repository which includes all information and resources about a library and its services [Fig.1].
- The system is remotely accessible across the Internet and/or a corporate intranet, support any hardware platform and run on any operating system.
- A friendly Web-based user interface which displays advice information and training resources in accordance with the user preferences and goals.
- The system runs independently from any software development environment and API.

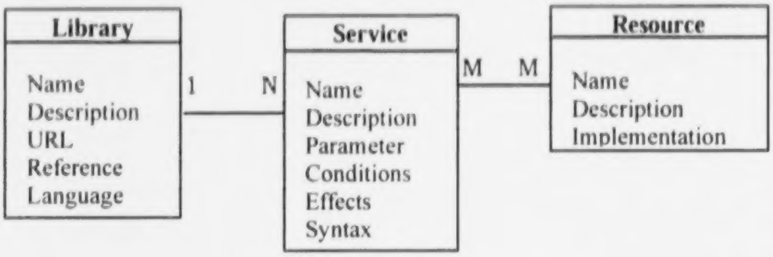


Figure 1. Main objects, their Attributes and Relationships between Objects

In the system, we make a distinction between two kinds of resource:

- Resources promoting understanding or dispensing further information. Examples of these units include HTML documents, videos and simulations. The learner exploits these resources to achieve a greater understanding of the domain knowledge.
- Resources describing problem-based learning activities, cases studies and demonstrations. These units enable the learner to attain a coherent and generally unique instructional objective among those specified in the curriculum.

Further work

The suggested system helps to reduce the training cost when a software library is first introduced. The system has also the potential to support the sharing of resources about services and libraries. It will also ease the transfer of advice-giving and intelligent training and advising systems to the real World. To achieve these objectives, the next step in our project aims to implement a Web based interface that enables developers to add new libraries, new services and new resources in the repository.

References

[Desmarais 93] Desmarais, M.C., Giroux, L., Larochelle, S. (1993). An advice-giving interface based on plan-recognition and user-knowledge assessment. *International Journal of Man-Machine Studies*. Academic Press Limited 901-924

[Kearsley 88] Kearsley, G. (1988). Online help systems: design and implementation. *Human Computer Interaction series*. Ablex Publishing Corporation

[Poulin 94] Poulin, S.J. Werkman, K.J. (1994). Software reuse libraries with mosaic. *Proceedings of the 2^nd International World Wide Web Conference*. Chicago, Illinois

[Rosson 93] Rosson, B., Carroll, J.M. (1993). Scaffolded examples for learning object-oriented design. Communications of the ACM. 39(4)

Silverman, B.J. (1992). Building expert critiquing systems: a situated tutoring alternative. *In proceeding of ITS'92* *conference*. Montreal (Canada).

[Brunner 96] Brunner, M., Kuhnapfel, B., Schroeder, U. (1996). New media in software engineering education. *Proceedings of the Third International Workshop on Software Engineering Education*. 18th International Conference on Software Engineering. Berlin (Germany).

Designers Designing a Web Site to Teach Designers to Design for the Web

Cary Staples
Department of Art
University of Tennessee . Knoxville
Knoxville . Tennessee . USA
staples@utk.edu

Jamison Abart
Visuoso Media
Knoxville . Tennessee . USA
abart@visuoso.com

Tim Parker
Visuoso Media
Knoxville . Tennessee . USA
abart@visuoso.com

Experience has shown us that designers are always interested in utilizing state-of-the-art technology, but despite the changes in technology, the basic principles of design do not change. Good design is about good communication, regardless of the medium. The process of integrating new technology into the existing practice of graphic design is an integral part of being a designer, as well as educating future designers and future clients.

The internet, an entirely electronic medium, creates new and different presentation issues, distinct from information that is organized to be presented using the print medium. As most designers have learned to evaluate information for print, this seems to be the most obvious place to begin to identify the issues that relate to design for the web. As the web is so new and inconsistent, there has not yet been a thoughtful investigation of the criteria necessary for effective presentation, resulting in enhanced communication. This combination of circumstances, the interest of the students to explore this new electronic medium and the newness of the medium itself, creates a unique environment for graphic design students in that they can participate in establishing professional criteria.

As with most graphic design programs, our computer lab has evolved out of the desktop publishing field. Previously, Art 405 was an advanced class about computer design for print issues. Today it has evolved into a digital media class where we teach design for CD ROM interfaces and, beginning this year, web design. As our curriculum cannot incorporate comprenhensive training in programming in addition to dealing with design theory, we approch the class from the perspective of Information Analysis. We analyze and evaluate information and determine whether the information should be presented as print, multi-media or CD ROM, or as a web site and discuss the merits and limitations of each platform.

For this project we tried to encourage the students to address issues that would be relevant to design and designers. We created the problem of how to encourage visual designers (practicing print professionals) to work on the web. Because the Web is such a dynamic media, many designers have not had the concentrated time necessary to become literate about web design restrictions. And when they do, the restrictions change. In our effort to teach the students about the web we used the web itself. In the process of learning to design for the web the students created a site that attempted to educate professional designers about design for the web.

Class projects and critique insights were posted to allow visitors access to the designer thought process. Projects include a site map in addition to a design brief. This was ment to help visitors see the difference between intention and reality. Concluding insights were also posted to allow viewers to understand the analysis aspect of the design process. Outside reviewers were invited to critique the student work and offer additional observations. We hope, in the future, critiques will be able to be carried out on line in a more dynamic environment.

Another feature that was built into each "project" on the site was a timer that would disclose the amount of time each designer has spent creating that particular site. This will help practicing designers, as well as prospective clients, estimate the amount of time necessary to generate a project.

Some of the design issues that arose in construction of the class site were:

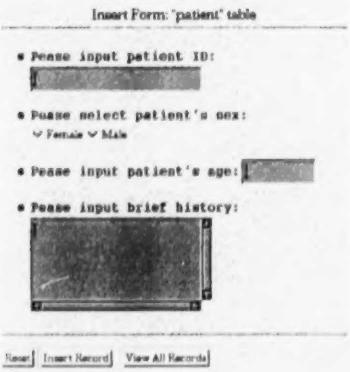
. . . the order of information presentation so that it would be most helpful to practicing designers and build upon

their previous experience. This would build confidence.

. . . the site design itself. We thought the design needed to reflect good design practices without being "over" designed.

. . . the fact that most of the users would be from a print background ment that they would print out the tutorials and would not be inclined to read on screen.

The site consists of assignments, tutorials and reviews. In short, the site is a mirror of the class and it evolved as the class evolved. We are making arrangements with the AIGA (American Institute of Graphic Artists) in New York to house the site and make it accessible to professional designers and design students worldwide. Our hope is that the site will remain as dynamic as the web itself.


Insert Form: "patient" table

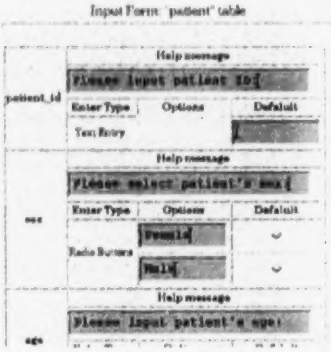
- Pease input patient ID:
- Pease select patient's sex:
  ○ Female ○ Male
- Pease input patient's age:
- Pease input brief history:

Figure 2: Input form

Figure 3(a): Design input form step1

The target specifies the value string, against which each record will be compared in answering the query.

The query form in Figure 4 was created with the help of the interface as shown in Figure 5. At this point, three query types are supported. First, it can display all the values of a given field for easy selection. Second, it can display all comparison operators together with an input value area. The user can select the appropriate comparison operator and then enter a text string as the query keyword. The third query type is used for specifying relational queries between tables. In this case, a user can select the tables to be queried. By default, the gateway (in Figure 4) also provides specification of display method for the query results: fields to be displayed and the sorting order of the records.

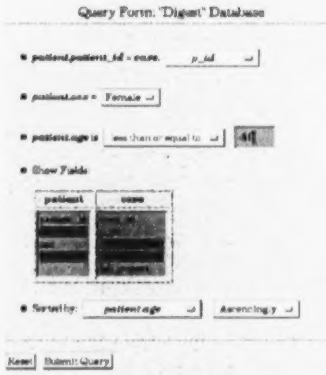

Input Form: "patient" table

patient_id
Help message: Please input patient ID:
Enter Type: Text Entry

sex
Help message: Please select patient's sex:
Enter Type: Radio Buttons
Options: Female, Male

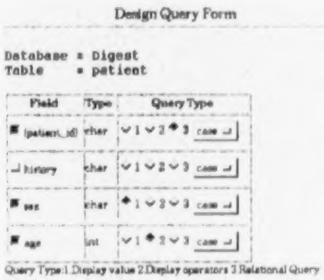
age
Help message: Please input patient's age:

Figure 3(b): Design input form step2


Query Form: "Digest" Database

- patient.patient_id = case. p_id
- patient.sex = Female
- patient.age is less than or equal to 40
- Show Fields
- Sorted by: patient.age Ascendingly

Figure 4: Query form


Design Query Form

Database = Digest
Table = patient

Field	Type	Query Type
■ patient_id	char	○ 1 ○ 2 ● 3 case
☐ history	char	○ 1 ○ 2 ○ 3 case
■ sex	char	● 1 ○ 2 ○ 3 case
■ age	int	○ 1 ● 2 ○ 3 case

Query Type:1 Display value 2 Display operators 3 Relational Query

Figure 5: Design query form

3. Conclusion

In this research, we have implemented a database gateway tool that has the capabilities for *information access and update, database design and creation, information output format, and automatic input and query form layout design.* By using standard web browsers as a uniform interface to all database transactions, it makes the database gateway adaptable to multiple platforms and reduces the need for using a customized client application for each database. Furthermore, we offer a progressive solution to database interface design. The form design tool allows the average users to automatically generate and modify HTML form documents for data entry and query, instead of directly changing the scripts that generate the corresponding hypertext documents. Such a tool supports improved maintenance of web databases as well as error-free and efficient interface design. As a result, even those who are not familiar with programming languages, HTML tags, or SQL queries can manage databases and design user interfaces. It can also reduce the time and cost of software development for database administration over the network.

References

[Fielding et al. 96] Fielding, R., & Frystyk, H., & Berners-Lee, T., & Gettys, J., & Mogul, J. C. (1996). Hypertext Transfer Protocol, 1996, Available at http://www.w3.org/pub/WWW/Protocols/HTTP/Issues/Revs/Rev81Clean.txt.

[McCool 94] McCool, R. (1994). The Common Gateway Interface, Available at http://hoohoo.ncsa.uiuc.edu/cgi/overview.html.

[Eichmann et al. 94] Eichmann, D., & McGregor, T., & Danley, D. (1994). Integration Structured Databases Into the Web: The MORE System, *Proceedings of the First international Conference on the World-Wide Web*.

[Bjorn and Hotaka 95] Bjorn, M., & Hotaka, R. (1995). A WWW Gateway for Interactive Relational Database Management, *Proceedings of The First Australian World-Wide Web Conference*.

BEST COPY AVAILABLE

Using Internet Technology and Performance Support Methodology for Teacher Training

Ginger Swope
C4610 University Center
Florida State University
Tallahassee, Florida 32306-4018
e-mail:gswope@cpt.fsu.edu

George Huffman
C4609 University Center
Florida State University
Tallahassee, Florida 32306-4018
e-mail:ghuffman@cpt.fsu.edu

Phyllis McVoy
C4613 University Center
Florida State University
Tallahassee, Florida 32306-4018
e-mail:pmcvoy@cpt.fsu.edu

Introduction

The Matrix of Services, a new state mandated funding model for Exceptional Student Education (ESE), will be implemented in approximately 3000 schools in August 1997. By that date, approximately 25,000 teachers and staff need to be able to complete the matrix form and understand the new concepts and terms it includes.

Mentor systems and inservice training are integral ingredients of the teacher culture. Current literature indicates, however, that transfer of learning from inservice training is limited. Concepts presented in traditional training activities are not internalized and embedded into memory and then utilized effectively during job performance. Peer mentoring has proven to be a highly effective method for increasing the rate of transfer. Two substantial drawbacks to mentoring are: (1) the mentor may not be available when guidance is needed and (2) in providing support, mentors are drawn away from their own jobs.

Transferring those familiar education models to the Internet is the challenge and opportunity of producing the Internet based Matrix of Services (hereinafter referred to as the Matrix Mentor™). Other indicators that traditional training methods would not be sufficient included: the introduction of a new work task (the Matrix of Services), a large and geographically dispersed audience, and the need for consistent training delivered in a short time frame.

Project Overview

The Matrix Mentor was designed using performance centered design concepts. Performance centered design is an iterative process which incorporates rapid prototyping with usability and performance testing. Performance Support Systems (PSS) typically provide users with the information, advice, and learning experiences they need to do their work. Training and support are provided at the moment of need. The Matrix Mentor includes the following performance support elements: online help, tutorials, and examples.

Goals

During the design requirements data gathering, the following goals of the Matrix Mentor were identified:

- Provide an alternative training model to the current inservice training
- Supplement the existing training
- Support teachers with the decisions about type and level of services to be provided
- Provide an online version of the Matrix of Services form

Additional goals that involve using the Matrix Mentor for more strategic purposes, may include helping instructional staff plan their delivery of services and/or becoming a tool that school districts and principals use to plan and provide resources indicated by the Matrix of Services.

Audience

The primary audience is anyone who completes the Matrix of Services. From the data gathering and evaluation material, we know that ESE teachers, specialists, coordinators, and directors, Speech/Language therapists, Student Services staff , and general education teachers are typically involved.

Functional Requirements

Functional requirements are the capabilities that a solution must provide, regardless of the type of solution. The following were identified as functional requirements:

- An electronic means by which to complete the form
- Clarification of terms / a glossary
- Varied approaches to accommodate different learning styles
- Context-sensitive support
- A simple method of calculating the Total Hours in School Week and Time with Non-ESE Peers

Internet Solution

Based on the results of the design requirements data gathering, we are developing an HTML based Matrix Mentor which will:

- Reach a majority of users because it is platform independent
- Provide at least a core of the functionality to all teachers
- Meet the functional requirements
- Be flexible enough to add functionality, especially given the fast pace of growth of Internet tools

The Matrix Mentor can not depend on a network connection because many of the computers in schools do not have an installed network. A future release may use a network connection to store and access data on a server.

The Matrix Mentor was written in HTML and JavaScript designed to be run using a web browser. Netscape Navigator 3.x was chosen as the browser because of its features, stability, and cross platform availability. The features include a more complete scripting language (JavaScript) which allowed for mouse over highlights on menus and the ability to set focus to a Netscape window (critical for bringing up Glossary windows, for example.) A Netscape plug-in was written to allow student data to be stored on the local hard drive.

The Matrix Mentor is delivered on a CD-ROM which includes video training and on diskettes without the video. Both methods include installations for Windows 3.1, Windows 95, and Macintosh computers with operating systems of 7.1 or higher.

Flora of North America: A Distributed Cognitive System

Kay L. Tomlinson
Center for Botanical Informatics, Missouri Botanical Garden, U.S.A.
Email: tomlinso@mobot.org

Mark A. Spasser
Center for Botanical Informatics, Missouri Botanical Garden, U.S.A.
Email: mspasser@cbi.mobot.org

John L. Schnase
Center for Botanical Informatics, Missouri Botanical Garden, U.S.A.
Email: schnase@mobot.org

The idea that a group of individuals will exhibit cognitive properties different from the properties of the individuals themselves is generally accepted [Hutchins 1996]. This phenomenon will have increasingly profound effects on human society as advanced technology creates opportunities for new groups and distributed, virtual communities to develop without reference to the physical location of the individuals involved. Finding appropriate situations to analyze, facilitate, and begin to understand the process of distributed cognition is therefore of great importance. The Flora of North America project (FNA), whose 850+ participating scientists make it one of the largest scientific collaborations currently funded by the National Science Foundation, provides such a test bed.

FNA was organized to combine the efforts of specialists in many different plant groups to produce a single 30-volume compendium of all naturally occurring plants in North America north of Mexico. Up until now, the project has operated as a traditional, paper-based publishing enterprise but is shifting to an electronic publishing format [Schnase et al. 1997]. The FNA information space consists of nomenclature, descriptions, distribution, ethnobotany, illustrations, etc., of more than 20,000 plant species distributed over more than half a continent. It is clearly beyond the grasp of any single individual, and cognitive tasks are distributed among members of the group. Our goal is to produce a system that will facilitate cooperative interactions and interchange of ideas among different parts of the project in a Web-based environment, the FNA Internet Information Service (FNA IIS), and thereby enhance the cognitive power of the community as a whole [Roberts 1964], [Hutchins 1996].

Details of the FNA publishing process have been given elsewhere [Schnase et al. 1997]. In brief, specialists are invited by the FNA Editorial Committee, the project's 35-member governing body, to prepare treatments describing various taxa; collections of taxonomic treatments, including distribution maps and illustrations, are then edited, reviewed, and assembled into printed volumes. Authors prepare "treatments" that provide data for the FNA database, electronic publications, and the printed flora. Treatments generally focus on species within a single genus, and each is prepared and reviewed by specialists. Authors study plants in the field, examine herbarium specimens, and review published reports of previous work. These functional activities are, for the most part, performed by geographically distributed referees and editors. At present, a mix of electronic and paper documents are used throughout.

As we set out to develop a Web-based computing environment for FNA, we realized that since no individual can comprehend the entire FNA ecosystem, it follows that there is no single *viewpoint* from which it can be conceptualized. This leads us to the idea of managing information by means of dynamically constructed activity-and-information spaces, or *role-based views.* Role-based views are derived from the socially constructed roles that exist within the project and represent the various information requirements, tasks, and responsibilities required for the FNA cognitive system to work. Taxon editors, for example, can be given the responsibility, and system functionality, to add a new author to the system or make an author/treatment assignment. Manuscripts being developed by each author can be viewed by the author; the editor responsible for that author and his or her treatments can view relevant information on the author's activities and associated treatments; editors view and manage the treatments assigned to them; and the project editor can view the entire enterprise. The various role-based views are delivered through dynamically constructed, personalized home pages that can be accessed using any current Web browser. A profile database maintains state information, and activities are implemented by a library of CGI scripts and simple form interfaces. Personalized pages are built in response to user logins.

FNA IIS accomplishes three things which are significant in terms of distributed cognition: First, FNA IIS lessens individual cognitive load by delegating information and task organizational duties to an external representational structure, i.e., the interface; it does the organization "behind the scenes" and presents to the user organized and encoded information that is in the right place at the right time and is easier to use. Second, FNA IIS enhances system performance by enabling massively parallel simultaneous use of a large information space via mapping of permissible views plus suites of operations onto individual and group knowledge resources and capabilities. [Hutchins 1996] reminds us that distributed cognitive systems achieve their information-processing power by superimposing several kinds of representations, or representational structures, on a single framework. In our case, the framework is a single, very large information space. Finally, FNA IIS instantiates role-based views over the information space. It structures not just the information but also the tasks: it simultaneously affords and constrains opportunities for the user to interact with the information.

To study multiple views of a cognitively distributed system like FNA, we will employ multiple data collection methods, such as server log analysis, protocol analysis, and unstructured interviewing methods. Methodological and data collection triangulation are complex, synthetic processes in which data derived from one method are analyzed in the context of those obtained by other methods [Janesick 1994]. For example, monitoring user-system interaction provides a check on whether users actually do what they say they do in interviews or when surveyed. Data will be analyzed using the grounded theory approach [Strauss & Corbin 1994].

Finally, we will use the FNA model to examine the implications of distributed cognitive systems for the practice of systematic biology and for biological informatics as a whole. At a time when biodiversity is declining faster than we can study it, enhancements to the cognitive capacity of science, such as may be facilitated by the Web, are crucial.

References

[Hutchins 1996] Hutchins, E. (1996). *Cognition in the Wild.* Cambridge, MA: The MIT Press.

[Janesick 1994] Janesick, V. J. (1994). The Dance of Qualitative Research Design: Metaphor, Methodolatry, and Meaning. In: Denzin, N. K. & Lincoln, Y. S. (eds.), *Handbook of Qualitative Research* (pp. 209-219), Thousand Oaks, CA: Sage.

[Roberts 1964] Roberts, J. (1964). The self-management of cultures. In: W. Goodenough (ed.), *Explorations in Cultural Anthropology: Essays in Honor of George Peter Murdock.* McGraw-Hill.

[Schnase et al. 1997] Schnase, J.L., Kama, D.L, Tomlinson, K.L., Sánchez, J.A., Cunnius, E.L., and Morin, N.R. 1997. The Flora of North America digital library: A case study in biodiversity database publishing. *Journal of Network and Computer Applications.* Vol. 20, 87-103.

[Strauss & Corbin 1994] Strauss, A. L. & Corbin, J. (1994). Grounded Theory Methodology. In: Denzin, N. K. & Lincoln, Y. S. (eds.), *Handbook of Qualitative Research* (pp. 273-285), Thousand Oaks, CA: Sage.

Acknowledgements

This work is supported in part by grants from the Andrew W. Mellon Foundation and the National Science Foundation (DEB-9505383).

WASE: Towards a Web-Aided Software Engineering

Ju An Wang
Faculty of Science and Technology
University of Macau, P.O.Box 3001, Macau
E-mail: fstjaw@wja.sftw.umac.mo

Since the early 1980s, hundreds of different CASE (Computer-Aided Software Engineering) tools have been developed. However, tool-making in software engineering is still in its infancy compared with other engineering disciplines. Several tools do exist for specific stages of software life cycle; however, these tools are not integrated together, nor have the benefits of cross-platform support. As a result, these tools can not be widely used by software development teams whose members distribute geographically and access different types of computing platforms.

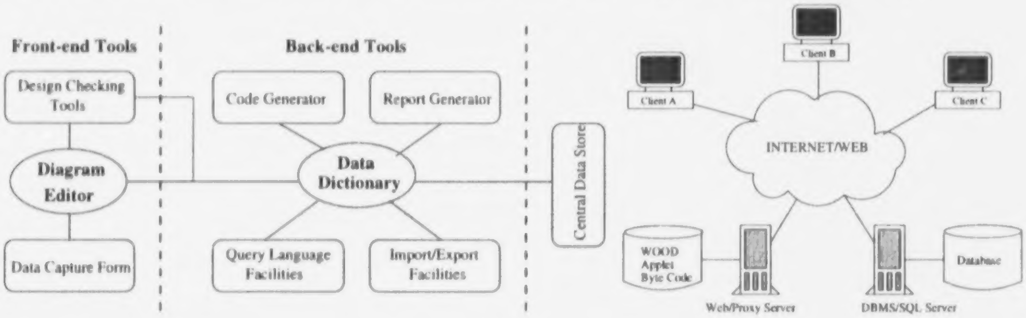
Today, we are living in a multi-platform world. That is why specific platforms must be chosen before we start our software development process. Such platforms serve as two purposes: the target platforms, which host the final products; and the developing platforms, which are used by developers during development. Usually the developing platforms should coincide with the target platforms. The availability of Internet and the World Wide Web (abbreviated as Web in the following) is changing the way of software development. The Internet is a typical distributed computing environment. Rooted from research collaboration, the Web provides a suitable collaborating platform for distributed software engineering environment. This suggests a new software developing paradigm: Web-Aided Software Engineering, written as WASE for short. Software engineers expect to work at home or at different locations, spanning geographical states, countries, or continents, with different platforms. How to handle the problems of managing complexity both in the product being developed and in the development process considering the developers distributed world-wide?

Needless to say, software developers need an environment in which all developing resources are widely accessible and several tools are available to achieve interoperability. WebSEE [6] is a testing WASE environment which takes the advantages of the Web as a universal interface and communication medium. Since all the tools here are implemented in Java, WebSEE is platform independent, providing high compatibility with users' existing computing environment, provided they have Java-enabled Web browsers installed. With the popularity of Java-enabled Web browsers such as HotJava, Netscape and Microsoft Internet Explorer, almost every computer is installed with one of such Web client software. As a consequence, WebSEE provides a cross platform toolkit for software development for virtually every Web user everywhere in the World.

It has been recognized that integrated tools are more useful and cost-effective than individual ones. WebSEE is an integrated tool-set which will support all five levels of integration proposed by Wasserman [4]: (1) platform integration; (2) Data integration; (3) presentation integration; (4) control integration; and (5) process integration. Take platform integration, for instance, which means that the tools run on the same hardware/operating system platform. With Java, CGI, JDBC and other Web technologies, WebSEE creates a seamless integration for software developers to share various developing resources based on Internet which is a heterogeneous network with different computers running different operating systems. Data integration refers to that different tools can exchange data among them. Thus results from one tool can be passed as inputs to another tool. WebSEE supports data integration especially at the level of shared repository. The tools are integrated around a centralized Web server which maintains a public, shared hypermedia object database describing the data entities and relationships among them. Different WebSEE tools can manipulate these data entities and modify their relationships. Presentation or user interface integration means that the tools in a system use a common metaphor or style and a set of common standards for user interaction. WebSEE chooses the Web browser as a uniform user interface so that different tools have a similar appearance. Users have a reduced learning overhead when a new tool is introduced as the Web provides hypermedia on-line help documents and training courses. WebSEE users can browse local information as well as remote relevant information, even chat with remote experts for consultation. Furthermore, it has been proved that the Web is a suitable presentation vehicle to support software engineering process due to its overwhelming popularity, low requirements and easy to use, graphical user interface, cross-platform support, dynamic display for spatial and continuous data, so on and so forth.

At present we have implemented in WebSEE several tools including requirements specification tools, training and communication tools, and software design tools. Let us concentrate on an object-oriented analysis and design toolkit, abbreviated as WOOD (Web-aided Object-Oriented Design), which is a collection of CASE tools integrated with the Web, supporting object-oriented analysis and design. For the current version of WOOD Version 1.0, it supports OTM and Catalysis object-oriented analysis and design methodologies. Other object-oriented software engineering tools will be supported by WOOD in the future.

WOOD adopts a client/server architecture to handle distributed computing requirement. Several clients are accessing WOOD through their own Web browsers. A Web Server would provide different instances of WOOD shell for different clients. As shown in the figure below, tools are basically divided into two main categories: the front-end tools and the back-end tools. The front-end tools are responsible for interfacing with the users to help *data creation, data gathering, data retrieving*, and *data updating*. On the other hand, the back-end tools are used for *data processing* and *data transforming*.



The **Diagram Editor** and **Data Dictionary** play important roles in WOOD. They are used to created object diagrams, structure charts and other design representations. Data captured by WOOD is stored into a normal relational database as *objects*. With the Java object serialization technique and Java DataBase Connectivity (JDBC), we established an object database through a traditional relational database. The Proxy Server is used to separate the Web Server with the Database Server. The actual connection from WOOD applets to the real Database Server would be transparent to the WOOD clients. The repository in WOOD is a centralized data store holding various types of Java objects. Concurrent access control is provided which maintains the data consistency and integrity.

The powerful toolkits in WebSEE will lead to improvements in productivity of software analysis and design taking full advantages of the Web environment. Details in designing different sub-systems of WebSEE are omitted due to space limitation.

References

[1] Ghezzi et al., Fundamentals of Software engineering,, Prentice-Hall, Englewood Cliffs , NJ, 1991.

[2] McDermid J A, Software Engineer's Reference Book, Butterworth-Heinemann, Oxford, 1991.

[3] Sommerville I., Software Engineering, 5th eds, Addison-Wesley Publishing Company Inc. 1996.

[4] Wasserman, A. I., Tool integration in software engineering environments, in Proc. Int. Workshop on Environments, Berlin, 1990.

[5] Ju An Wang, The Web: Promise and Challenge for Software Professionals, in Proceedings of Macau IT Congress'97, pp 101— 106, January 1997.

[6] Ju An Wang, WebSEE: Towards a Web-based Software Engineering Environment, Technical Report FSTSFTW TR961011, University of Macau, June 1996.

Acknowledgments The author would like to thank Java Working Group, consisting of Jack Wu, William Lin, Willian Kuok, and Alica To, for their work on WOOD implementation.

BEST COPY AVAILABLE

ROADMAP TO ATM

Glenn Warnock, Professor, Computer Studies Dept.
warnocg@algonquinc.on.ca

Douglas Reid, Professor, Electronics Dept.
reidd@algonquinc.on.ca

Maike Miller, Professor, Electronics Dept.
millerm@algonquinc.on.ca
School of Technology and Trades, Algonquin College
200 Lees Ave., Ottawa, Ontario, K1A 0C5

This paper presents the objectives and accomplishments of the 'Roadmap to ATM' project. The project is a collaborative effort of Algonquin College , Communications Research Centre of Industry Canada (CRC), Knowledge Connections Corporation (KCC) and the Ottawa-Carleton Research Institute (OCRInet) with Algonquin College being the lead agency. The first stage was completed on May 31, 1997 at a development cost of approximately $400,000.

Objectives

The project objectives were to:

- demonstrate the full potential and effectiveness of an ATM network in delivering media-rich courseware to on-line learners.
- develop a network-based course delivery strategy that supports a high level of student-teacher interaction.
- deliver the video components of the learnware over a native ATM network.
- provide reports on the production process, the delivery strategy and the effectiveness of the course.

Project Description

The 'Roadmap to ATM' project consisted of developing, delivering and evaluating a pilot course to provide an overview of Asynchronous Transfer Mode (ATM) networking technology. It has the approximate content equivalent of a five day classroom session. The pilot delivery of a preliminary version took place in May of 1997. Work is currently in progress on refinement on this version in preparation for delivery over a high bandwidth TCP/IP or ATM network.

One focus of the project is to demonstrate and evaluate combinations of teaching methods, learning models, multimedia development applications, delivery mechanisms, and supporting technologies for effective distance learning and on-line learning given the availability of a broadband network for delivery and server based learnware. The learnware will be kept up-to-date as a living document incorporating the work (from assignments and group projects) of the students into the database. Thus future students can build on the accomplishments of their predecessors.

The Roadmap to ATM is comprised of several components. The core of the system is similar to traditional computer based courseware with extensive use of graphics, audio and video components and animated simulations. The Common Room area provides a facility for students to post questions or comments and for the professor or other students to respond. It also contains reference materials and links to other relevant reference areas. The video conference capability allows for real-time "lecture sessions" with the professor, other experts or on-line student-teacher conferences.

The system is built on standard World Wide Web technology. The courseware is developed in Authorware, converted to the Shockwave format and then delivered over the network by a standard Apache HTTP server. Asynchronous conferencing (the Common Room) is implemented using the Hypernews CGI. Video conferencing is provided through standard MBONE tools. OCRInet is an ATM network linking research institutions, educational institutions and telecommunications companies in the Ottawa area. It provides a minimum of DS-3 (45 Mbps) connections between all parties. The clients run Windows NT and the server runs Solaris 2.5 (UNIX). The server's 8GB of disk space determined the amount of video to be incorporated.

Results

The Roadmap to ATM pilot was given a high rating by the evaluating students. They gave a good rating to the quality of the course material, the graphics and the audio and video components. They appreciated the availability of varying technical levels in the course. Recommendations were made for improvements to the user interface and navigation through the courseware. Requests were made for more simulations. The ability to add and modify the course content and have it instantly available was very valuable.

Although the pilot was intended to be run from four designated NT workstations with ATM connections to the OCRInet, a group from Telesat Canada expressed an interest in participating in the pilot. They were able to set up a Windows95 client with an Ethernet connection to the OCRInet and participate without any difficulty.

Authorware is a very effective tool for the development of this type of courseware. There is a moderate investment required in learning the tool, especially the more advanced features and the programming interface. Shockwave works well to provide a Web-based format of the Authorware content. It provides the further advantage of some protection from copying of the original content (especially expensive graphics) by converting the content to a run-time format. The greatest difficulties involve the management of the large number of files produced when large Authorware modules are converted to the Shockwave format.

Many obstacles were encountered in the development of the native ATM video streaming component. The ability to stream video files between the UNIX server and the NT client was demonstrated, but there was insufficient time to integrate this component in the system. It was also found to be too expensive to encode the video as MPEG-2 files as originally intended, so MPEG-1 encoding was used instead. The encoded video was then incorporated directly into the Authorware materials instead of being delivered separately.

The interactive components were not developed and exercised to the extent desired. Although the asynchronous conferencing was available for the pilot, it was not used much because of the short duration of the sessions and the fact that the students were not widely dispersed geographically. Problems were encountered with the MBONE conferencing tools and with multicast routing through the network. As a result it was not possible to use video conferencing to run the lecture or real-time conferencing sessions.

Overall the project demonstrates the potential of higher speed networks to deliver media-rich courseware with an interactive component. However, the effort required in the production of high quality courseware materials must not be underestimated. In addition, more capable tools are required for the organization, management and enhancement of the on-line course material, particularly as the volume increases with additional courses.

Although the potential of a network for facilitating student-teacher and student-student interactions was not demonstrated in this project, it is well known. Certainly the development of the Roadmap to ATM course itself would have been far more difficult without a high-speed network to support the collaboration of widely separated participants. Perhaps the final assessment of the project should be made on the basis that most of the students involved in the project were immediately hired in the local telecommunications industry.

A Distance and Open Learning Project for Multimedia Training

Gérard. Weidenfeld Université de Picardie Jules Verne - DESS SIM - IAE - France
Gerard.Weidenfeld@u-picardie.fr

Introduction

Enclosed herewith is a description of a distance and open learning project related to a postgraduate course, the DESS « systèmes d'information multimedia ». The contents of this course are from one hand related to basic computer topics such as object modeling, knowledge representations, computer networks, human-computer interactions, computer graphics, hypermedia and indexing approaches and from the other hand related to methodology and practice in developing multimedia applications. The participants of this course are graduate students and computer technology engineers but many potential trainees cannot attend a course of four hundred hours for geographical and professional reasons. This is the starting reason of the actual project.

One of the main characteristics of the DESS course is to emphasise both theoretical and practical abilities. In the actual organisation the development of an application is central and yields various activities : need and market analysis, comparision with related solutions, adaptation of design and development related experiences, adaptation of evaluation models, etc. Aside of these activities, students are required to make inquiries, to build products catalogs or to make bibliographical researches. Intensive use of information is common to all these activities and the net is the « canonical » channel for this information. The distance learning approach will be to provide the student with computer tools : multimedia *courseware* for acquiring knowledge and an *activity platform* for acquiring know-how and for applying knowledge contextually. In this process most of the communication between students and teachers is performed through electronic connection, especially E-mail.

Courseware provided to the student

The first stage was the realisation of a book [Wei 97] and the design of an associated hypermedia with :

Automatic generation of standard links by means of formal structure [Smi 92].

Adaptation of approaches used in documentation [Coo 93] : a typology enables to classify the indexed terms and define different frames of cards.

Generation of semantic links which are either derived from a thesaurus [Flu 97] or are « meta knowledge » which is obtained by pointing out relationships between contribution of the various authors [Wei 89].

The interaction provided in this hypermedia is somehow too poor to stimulate the learners. The following features are intended to put the student in problem solving contexts and to favour interaction. Four interactions (hypermedia, simulation, animation and access to computer tools) will be used in order to illustrate the computer topics used in multimedia : object analysis tools and databases, interface builders, programming languages, image processors, geographical information systems,...

The courseware offers a theoretical frame on which plugs wide information, activities and programs. The pedagogical methods used are problem based learning and use of simulation for professional training [Lee 96]. The courseware dos not include tutoring modules, these functions are held by the (distant) tutor.

Activity platform

The purpose of the course is not purely theoretical : more thant half of the time is devoted to the realisation of a project and évaluation of this project weights for nearly 50 % of the final examination note. The objective of the activity platform is to allow the students to train themselves to perform multimedia projects. For that purpose three kinds of tools and guided activities are proposed : a methodological frame presented through hypermedia interaction, a set of projects as commented examples and facilities to access the internet information base.

The important topics for the methodological frame are need analysis, selection of information, share of work to the various actors of a project, management of costs and delays, development methods and validation methods. The approach of the previous chapter is still applied here. The first step is a book [CLW 97] which hypermedia version will be derived in the way previously explained.

Examples of multimedia projects are choosen among those developped during the last four years by the students of the DESS SIM according to the proposed methodology. Among the numerous mock ups, those chosen highlight various aspects of project development : variety of objectives and domains, teams involved, etc. Of course results such as human-computer interface, graphical design, knowledge representation which determines the form of message coming from the computer reflects somehow the story of the project. Correlations and deviations between expected issues and reality provides also a powerful way to study that cases. Projects are related to courseware (assistance in medical diagnosis, accountacy, simulation for marchandising ([LeW 96])), cultural and entertainement and information bases and servers.

All these projects have been developped in the university so that a memory of these projects progress has been kept till now. This is the main reason why we have choosen to illustrate the methodological frame by means of these applications. Nevertheless added information will be taken from other realisations : commercial ones for which some innovative facts have been publicly reported or other projects.

Information about multimedia available through internet are available under a lot of viewpoints which are fundamental for any professional in the multimedia field. First of all, the various experiences and opinions of users will offer a view on multimedia applications. This will give major information concerning proeminent domains for multimedia, performed activities, demands of users to improve the interfaces, expected added value of multimedia and way to evaluate it,.ect. An other significant contribution is given by experiences and opinions of actual or potential actors of the multimedia field : hardware and software providers, graphists, film actors and directors, photographs, economists, ...

Actually, students are aware of the importance of information and some general principles are given to them for optimising the time spend in browsing on internet network. But, even for postgraduate students, efficient browsing is a hard issue and the main difficulty is to start with the *good request.* Definition of general models for « managing immense storage » is an old, difficult and on going research and development preoccupation ([Bus 45], [Nel 88], [Flu 97]). We are now designing « pedagogical » interfaces in order to help the student to formulate requests relative to dedicated activities (market analysis, need studies,...).

Perspectives

The action described here is still under progress and the first objective is to set up the distance learning course. During the next academic year availability of learning material will enable to start an experience with a significant amount (over 30 %) of distance interaction. During the 1998-1999 the final version will be held, with some experimental and evaluation features. Evaluation modalities are now set in details but they will take care of how the students and the trainers will use the overall devices.

An other direction is to adapt the described method to other contents. Business managers make a so intensive use of information that they need to acquire information retrieval abilities [Wei 95]. We are actually looking for an adaptation of the system described here in order to fulfill these needs.

References

[Bus 45] Bush V, *As we may think*, Atlantic monthly, july 1945.

[CLW 97] Cochard G-M, Leclet D., Weidenfeld G. *Conduite de projets multimedia, to appear at Masson, Paris, 1997.*

[Coo 93] Cooke P., Williams I., *Design issues in large hypertext systems for technical documentation*, in hypertext, theory into practice, Ray McAleese, intellect, oxford, 1993,

[Flu 97] C. Fluhr, *Techniques d'indexation et applications*, « Techniques de base pour le multimedia », Masson, Paris 1997

[Lec 96] D Leclet , G. Weidenfeld, *Building simulation for training in professional contexts*, UES 96

[LeW 96]D Leclet, G Weidenfeld, "*un modèle de simulation* " ITS 96 juin - Montréal. Springer 96

[Nel 88] Nelson T.H. *Managing immense storage . Byte, january 1988*

[Smi 92] J.M. Smith, *SGML and related standards*, Ellis Horwood series, 1992

[Wei 89]G. Weidenfeld, *hypermedia and tutoring;* Delta multimedia journal, octobre 1989, Editor C. Cazala & P. Held

[Wei 95] G. Weidenfeld,« *Exploratory learning and simulation games applied to management training* » in Telematics for Education and training, Editor P Held and W Kugemann, IOS Press,1995

[Wei 97] G. Weidenfeld et alii, « Techniques de base pour le multimedia », Masson, Paris 1997

Acknowledgements

This work is undertaken with financial support of the French Ministry of Education and of Regional authorities (Educapôle). It benefits from the contribution of my colleagues Gérard-Michel Cochard and Dominique Leclet , the publishing house, Masson and the French distance education organism, the CNED.

The Effects of Waiting Time on Website Utilization

Bruce D. Weinberg
Boston University
School of Management
Marketing Department
United States
celtics@bu.edu

The amount of waiting time between indicating a request to utilize (i.e., link to) and the subsequent availability of a webpage seems to be of significant importance to World Wide Web users and hence, of significant interest to marketers who use or plan to use the Web for marketing activities. This has become more noticeable as 1) the growth in the number of individuals using the Web seems to have outpaced realized improvements (i.e., what has been developed and *actually* implemented, installed or distributed) in network (e.g., pipeline width) or network-related (e.g., compression) capabilities, and 2) the amount of information available at a website seems to be increasing monotonically (i.e., websites seem to be providing more as well as richer -- graphics, sound and video -- information).

The waiting duration for a completed connection to a webpage results initially in two consumer behavior outcomes. A consumer either waits for the entire duration and has an opportunity to utilize (to some extent) a webpage, or does not wait for the entire duration and hence, does not utilize a webpage (e.g., the connection process is interrupted by clicking the "stop" button on the browser or by redirecting the browser to another address). It is clear that marketers want to minimize the occurrence of the latter outcome as it nullifies an opportunity for them to communicate with and to serve their current or potential customers (e.g., provide useful information, provide a product that best satisfies a consumer's needs).

A consumer who requests a connection to a website (URL address) and then chooses to abandon the wait at one point in time may be less likely to seek out making a connection at a future point in time; and as this type of behavior persists (i.e., a consumer attempting again to make a connection, and again choosing to not wait for the duration), the likelihood of seeking out a connection to a website may subsequently decrease. A consumer who requests a connection to a website and subsequently does not wait for the duration will be referred to as a "Nonvisitor."

A consumer who waits the entire duration and connects to a website, then chooses the extent of utilization of the current webpage (The term website is used at this point in hopes of no loss in generality when discussing waiting time with respect to webpage utilization. It is assumed that upon any connection to a website, a consumer will encounter a webpage associated with that website.). The consumer then decides whether to a) exit the website, b) utilize the webpage (i.e., read and process contained information) for some length of time (where the length of time is, in spirit, greater than zero seconds), or c) seek out another webpage within the website (the theory developed may be generalizable to webpages associated with a website, for example, links to other websites). This decision is assumed to be made during the entire time that one is connected to a particular webpage. A consumer who waits the entire duration and connects to a website will be referred to as a "Visitor."

Marketer objectives can be specified for each type of consumer, the Nonvisitor and the Visitor. The ultimate objective with respect to the Nonvisitor is that this type of consumer visit the website. A lesser, but directly related, objective is increasing the probability that this type of consumer visits the website. The ultimate objective with respect to the Visitor is that this type of consumer inspect all useful webpages available at the website (useful is a function of an individual consumer's informational needs, which in turn may be a function of its stage in the consumer decision process.). A lesser but directly related objective is increasing the probability that this type of consumer will visit a useful webpage. The degree to which marketers achieve these objectives could affect consumers' attitudes toward their website, the manufacturer, and the brand; and in turn the firm's revenues and profitability.

This research focuses on the effects of waiting duration on website utilization and draws on theories from marketing, psychology and economics that are related to perceptions of time (including waiting time), attitudes, and customer satisfaction. Strategies for reducing a consumer's perceived length of a waiting duration are discussed with the purpose of assisting marketers (as well as other website and webpage design interested professionals) in increasing consumer utilization of their websites.

In order to minimize perceived waiting time, and hence increase the likelihood that a consumer will (wait to) view all of a website's information, the theory suggests that website information should be allocated to individual webpages that are viewed in a specific order such that the waiting time associated with each webpage is in ascending order. That is, the waiting duration for the first webpage viewed by a consumer is shortest; the waiting duration for the second webpage viewed by a consumer is second shortest;...the waiting duration for the last webpage viewed by a consumer is longest.

Professionals involved in the process of designing a website need to consider not only the information to be presented, but also the waiting time consequences associated with this information. A website design process that places significant importance on the waiting time associated with information is more likely to result in a website that is satisfying to consumers and the firm. It is recommended that professionals engaged in designing a website carefully consider waiting time during this process.

Education on the Net: The Experience of the Writers in Electronic Residence Program

Herbert H. Wideman, Centre for the Study of Computers in Education, Faculty of Education, York University, Canada (herb@yorku.ca)

An ongoing five year study , "National Networks for Learning: Building Collaborative Inquiries", is investigating how two national telelearning networks are being implemented in the classroom. Qualitative methods are being employed to illuminate how these initiatives are being adapted and modified in different contexts, how they effect the day to day life of the classroom, and how participation impacts both students and teachers. The specific questions we are addressing include the following:

> Does participation in telelearning projects lead to a shift in teacher and/or student self-perception?
>
> Do they promote a change in a teacher's sense of what is possible in the classroom, and of what constitutes good practice?
>
> How do students perceive the innovation, and what effect (if any) does it have on their subject-related knowledge and skills?

This paper some initial findings from one area of this research, a study of student and teacher experiences of a nationally distributed, Internet based learning network that has been running in Canada for about a decade now: Writers in Electronic Residence (WIER). WIER was chosen for study because it is a relatively large network by Canadian standards, involving the participation of some 70 schools in any given year from all areas of the country and students in grades ranging from the junior elementary to the senior high school levels. WIER uses a network conferencing system - first class - to link writing and language arts students to Canadian authors, teachers, and each other for the exchange and discussion of original work. The authors, nationally known literary figures in Canada such as Kevin Major and Susan Musgrave, read student compositions that are sent to them by participating classes and send responses to each student, commenting on their work and sometimes suggesting revisions. The works are typically poems, short stories, or segments of a longer fictional work, which students draft and then submit to their teacher for uploading to the conference. A primary goal of WEIR is to facilitate student engagement in ongoing, reflective discussions about their posted work both in response to the professional writers (mentors) comments and to responses received from other students who have read their stories or poems.

Results

The data on student and teacher experiences with the WIER program being discussed here was gathered in face to face interviews with participants at 10 schools from across the country. The results of our preliminary analysis of the transcripts are presented from three theoretical perspectives: first, a consideration of the role of authentic audiences in promoting writing ; second, a look at how students come to develop a critical perspective on their compositions through WIER; and finally an analysis of the operational logistics of WIER participation and its impact on teachers.

Audience and Authenticity

All of the teachers and nearly all of the students placed a high value on the responses provided by the WIER authors to student work. Students greatly appreciated getting comments from a "real" writer who has published books to his or her credit, someone "who knows what he's talking about". They usually commented that the author's feedback is different from (and more useful than) that given by their teacher, that it is more comprehensive and deals with more fundamental creative issues such as character development, story structure, or the quality of description. But beyond the specifics of the response received, the student-author exchange clearly has a positive impact on students' self esteem that reveals itself indirectly in many student comments. They expressed surprise and delight that an author would read their work and take it seriously, and often indicated that they work harder at compositions that they intend to post to the authors. Gradually their sense of the value of writing as a rewarding vehicle for self expression began to expand. Having a real and valued audience moved those initially unenthused about writing away from a view of it as just another classroom chore. Teachers saw this in their students and cited a shift in student perceptions of writing and increased intrinsic motivation as a key benefit of WIER.

Development of a Critical Perspective

Students indicated that the authors' comments would often open their eyes to limitations or problems in their work that they had not been aware of. There were times when they would disagree with the writers' remarks, but in the great majority of cases they would see that what they authors had to say "made sense". While very few students would revise their posted story, most claimed that they applied the suggestions consciously to their next creative efforts, monitoring their work more closely to see if the earlier cited weaknesses had been eliminated from the new composition and/or the suggestions incorporated. When asked, most students felt that this process was increasing their ability to view their own efforts with a critical eye. This suggests that these students are beginning to internalize a more mature set of self-monitoring skills that should improve their work. It will be interesting to see if the textual analysis can offer some corroboration for these self-reports.

Operational Logistics

The operation of a WIER project by a classroom teacher without outside support is extremely time consuming, requiring several additional hours of work every week to upload and download compositions, monitor salons, and coordinate and administer WIER related activities. It was only because they so greatly valued the benefits they felt the program offered their students that these teachers were willing to put in the extra hours it necessitated. Because it was so demanding, it became a major (usually the major) component of the writing curriculum during its ten week run in the classroom. A few teachers tried different ways to lessen the workload, either by having an assistant or a student do the uploading or downloading/printing of stories, but this was only marginally helpful. If programs of this type are to expand beyond a self-starting group of early adopters, it will be necessary to reduce the operational drudgery involved in accessing and contributing resources to a central data pool via the Internet.

Discussion

Our initial analysis of student experience suggests that the WIER program meets many of the criteria called for in what John Wallinsky has termed the New Literacy [Willinsky 90]. By providing an authentic audience for writing, it fosters a literacy that arises from communicative acts rather than private development, and promotes a stronger sense of agency and identity as a writer in students; and through dialogs between student and mentor, it promotes a decentering, a move away from a naive writing stance to one that considers others' perspectives in critically reflecting on one's own work. What remains to be seen is whether these changes in students are sustained after leaving WIER, and whether they are reflected in the writing itself. Both of these questions will be addressed in the next phase of our work. We plan to examine the extent to which the specific recommendations of the author mentors which have general applicability are carried forward by students into their later compositions, and to study the longer term impact of the WIER experience on the writing of a sample of students over the next few years.

Reference

[Willinsky 90] Willinsky, J. (1990) *The new literacy*. New York: Routledge.

Overcoming Electronic Course Delivery's Greatest Obstacle: Specific Policy Recommendations for Institutions of Higher Learning

David Wiley, Marshall University
Office of Information Technology

Context

Advances in technology continue to increase our capacity to communicate greater quantities of information to our students in a manner which both increases their chances of learning and makes more efficient use of their time. Recent advances have even removed physical barriers of time and space, allowing students to acquire skills and knowledge even while temporally separated from their instructors. The purpose of this paper is not to review research into the effectiveness of distance learning technologies. However, this paper does assume that distance education technologies are effective methods of instruction.

In an educational marketplace which is becoming increasingly competitive, a university's ability to eliminate students' barriers-to-entry will predict its long-term fiscal viability. So-called "distance education" technology has overcome several of the physical barriers already. Higher education itself has been relatively successful in overcoming financial barriers by securing funding for the acquisition of "distance education" enabling technology. Faculty around the world are mastering distance education techniques and strategies, and producing online course content. So why isn't electronic course delivery taking off? Policy. Even with hardware, software, bandwidth, experienced teachers and completed courses ready for the offering, nothing can happen until higher education establishes policies that can govern this new medium.

Few people want to make mistakes. And even fewer want to make them in public. Because this is true, the vast majority of institutions of higher learning are "standing around," waiting for other institutions to implement electronic course delivery policies and work the bugs out. They're waiting to see what mistakes are made so that they don't make them themselves. Accordingly, the result is a lethargic motion in the direction of implementing "distance education" technology. Colleges and universities are, as it were, gingerly dipping their toes in the pool waiting for someone else to jump in and tell them "the water's fine."

Perhaps the best example of this hesitance is the issues of intellectual property and faculty compensation. Perhaps more than any other policy issues, these stand firmly between the new technology and the students. Because they seem so complex, and no one wants to make a mistake in implementation, efforts to create policy simply die in committee, and because ownership and funding issues don't get worked out very few online courses are offered. Almost all of the online offerings at institutions of higher learning around the world exist for one of two reasons: either the responsible faculty member received a grant or is highly self-motivated and forward-thinking. The universities themselves are doing very little to promote online delivery of course materials (with the exception of asking faculty to do extra work for free, or at best provide grant writing support) because they refuse to deal with issues of policy. Eventually universities must realize that as the number of methods by which a potential student can obtain knowledge and skills increases, and as the number of students in each freshman class decreases, the university must proactively compete, with both enthusiasm and creativity, if it will stand a chance of survival.

There are several steps to the creation of a successful electronic course policy. The first key is developing a long-term vision for electronic courses at your university and a flexible strategy of how to bring it about, or in other words, the much quoted "begin with the end in mind." Without a clear vision of where you are going, intermediate negotiations and decisions will be at best disjunct and at worst random and haphazard.

The second step is securing the involvement in the policy creation process of decision makers from each administrative department, and instilling the vision in them. These key administrators are either your greatest assets or worst enemies. The single greatest barrier to the creation of policy continues to be administrative inertia, the property of higher education described by the statement "we've always done it this way." If you gain the support of administrators with the ability to make decisions and commitments and follow through on them, the policy creation process will be significantly easier.

The third and perhaps most important step is outlining the current processes at work, and modeling their electronic equivalents to be as similar as is both possible and efficient over the long term. For example, trace the paper trail a student must traverse in order to register for classes. Then reproduce the process for electronic course students, carefully balancing administrative structure already in place against opportunity for increased efficiency presented by the new technology. Most universities will be unable to invent policies or procedures which are completely new and independent of existing ones for legal reasons if not for any others. Making as much use as possible of existing policy and administrative structure is the main key to success.

The fourth step is obtaining faculty feedback and getting faculty participation in creating the policies which affect them. This is a matter of basic democratic process and simple courtesy. Of course, faculty will abide more happily by policy guidelines which they help establish and feel some degree of ownership over. For issues such as intellectual property and compensation, getting faculty to participate should not be difficult .

The final step is determining guidelines by which the new policies will be reviewed, and the timeline for the review.

Specific Recommendations

At Marshall University, the issue of intellectual property / ownership of the new courses was the portion of the electronic course policy which caused the greatest stress. The E-course policy committee which had been established by the President created a draft document and presented it to the faculty of the university as a "request for comment." This provided a starting point for what would turn out to be negotiations so intense and heated they would have served as good practice for diplomats on missions to the Middle East. The discussion quickly drew most involved to one of two sides, those "representing the institution" and those "representing the faculty." A few committee members who "wear both hats" attempted to mediate, and the end result drew from policies already in place, research into current practice at other institutions, and appropriate tweaking. The main sticking point was whether faculty should be paid to develop electronic courses *and* maintain full ownership of the material. Faculty maintained that "works of art" such as photographs or works of music are commissioned and yet remain the property of the creator, and that course material should be treated the same way. The institution maintained that a faculty member who was paid $20,000 to develop several online courses who retained ownership of the material could leave the institution the next semester and take the courses with them, leaving the university out $20,000 with no courses to show for the students' money. The policy as finally approved by the President states that faculty maintain ownership of the material and the right to market the courses privately for profit, but the university has the right to use, free of charge, all courses whose development was supported by the university. In this way, both sides were able to get what they were really after: faculty "own" their courses and can take them with them if they move to another school, and the university retains the right to use free of charge the last version of a course whose development it supported. There are more details which are stated specifically in the policy, but this general arrangement is certainly a model which other institutions of higher learning will be able to use as the basis for successful policy creation.

Marshall University's policy governing all aspects of online courses will soon be available via the world wide web. For information about the document's location, see http://www.davidwiley.com/webnet/

SHORT PAPERS
(Works In Progress)

SUPPORTING COLLABORATIVE WORK IN INTRANET USING ACTIVE DBMS IMPLEMENTED WITH JAVA

ANIORTÉ Philippe, aniorte@iutbay.univ-pau.fr
DALMAU Marc, dalmau@ iutbay.univ-pau.fr
ROOSE Philippe, roose@ iutbay.univ-pau.fr
Laboratoire d'Informatique Appliquée
IUT Informatique de Bayonne
Place Paul BERT
64100 BAYONNE
FRANCE

Presently, the Intranet is more developped and its utilisation has increased. Companies such as Ford, Hewlett Packard and Silicon Graphics use about tens of internal web servers for Intranet. A current estimation shows that 70% of Netscape sales has resulted from Intranet sites.

Reasons for the rapid expansion of Intranet sites have been analysed in [Sand 1996]. On one hand, the basic technology (TCP/IP, HTML, ...) is easy to use and not too expansive. On the other hand, the Intranet is seen as a way for :

- implementing the new models of the entreprise where the organisation and staff must work together - theory of cooperative work,
- sharing the company information from the data bases,
- making the interface easy and independant from the platform

As well as the human dimension, it is commonly estimated that the applications of cooperative work in the Intranet includes the dimension of Information System (I.S.) and that the cooperation relies on sharing the company information from databases.

However, the actual DataBase Management Systems (DBMS) show their limits for these types of applications. Indeed, some studies which were carried out about this problem deplore the passive nature of these systems which only obey explicit demands - called requests in DB languages - from the user or the applications. In fact such systems are unable to automatically and instantaniously respond to events at the time of their occurence, which results in the user regularly consulting the base to be kept informed. This problem becomes more important when this DB constitutes the federating element of a cooperative work.

The response which we give to this problem naturally consists of using active DBMS which, by nature, allows to correct the problem in question. However, if the problems in the active DBMS research's area are well established, they must be looked into again for the management of cooperative work in the Intranet. It is one of the objectives of our contribution.

To do this, we took advantages of the experience gained during the conception of the active DBMS *ADACTIF* [Tawbi 1996] validated by an operational not networked prototype. In this prototype, the implementation of Event-Condition-Action (ECA) rules which classically describe the behaviour of the active DBMS is inspired by the Ada language and from its mechanism of parallel tasks and from the synchronization by *rendez-vous*. So when an event occurs, the *rendez-vous* is taken with a task which verifies that the condition is true in order to carry out the procedural action. The same principle of tasks and *rendez-vous*, used with the component events, allows the detection of complex events for which the algorythm of composition is also defined in a procedural way [Tawbi et al.1995].

Our active DBMS for the cooperative work on the Intranet conserves the same characteristics and originalities. Moreover, it is extented with the goal to manage the cooperant modules distributed on different sites linked by the Intranet and organised in groups of work whose constitution can evolue over time. These modules must be able to share the information in the database, to exchange messages and to circulate events on the network to implement the cooperative work. The schema below represents 3 places supporting 12 modules constituing 3 groups and one isolated module :

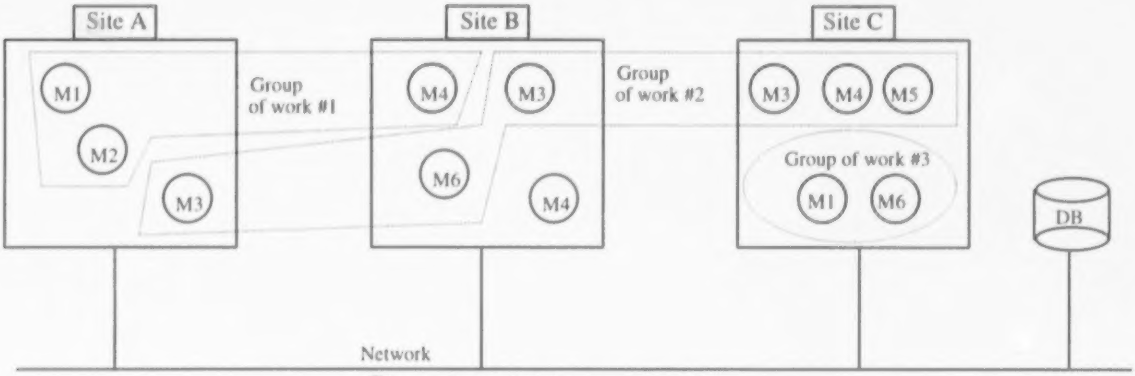


Figure 1 : Multi-place groupwork

Each place on the network has a Communication Manager (CM) wich allows to access the DB, to communicate with other places using a mail system and to take rendez-vous with distant rules. When an event occurs on a place, it is detected by the correspondent Event Manager (EM). The events the system needs to detect had been defined by the user. This definition includes a qualification which indicates where the event will occur and where it will be broadcasted (in terms of modules and groups of work). The EM uses this specification to transmit the event to others places and modules which need to be informed. On each place, this event is transmitted to the Rules Manager (RM) which will take rendez-vous with concerned rules.

These rules, implement the behaviour of active DBMS, using the ECA schema and carry out the action. Actions can be DB requests, any local operations or network based processing. In this case, the CM is used to start distant work (DB request on a place which doesn't have the DB, message transmission...).

Our system relies of the Java language. The advantages of this language can be found in its multi-platform aspects, communication and parallelism (Threads), on which our work is based. A prototype lets us validate the fundaments of our propositions. It uses parallel tasks able to communicate between them, with which the events can take *rendez*-vous through the network.

The perspective of exploiting our propositions in the domain of organization and management of a health system including the different services of an hospital and external services (general practitioner, convalescent home, ...) should supply us with a field of interesting experiments which will allow to make a critical analysis on our work to refine our propositions.

[Sand 1996] Sandoval V. (196). *Intranet le réseau d'entreprise*. Editions Hermès, 1996

[Tawbi et al. 1995] Tawbi C., Jaber G, and Dalmau M. (1995). Activity Specification Using Rendezvous. *Second International Workshop on Rules in Database Systems RIDS'95*. Springer, Lecture Notes in Computer Science, Vol 985. Athens, September 1995.

[Tawbi 1996] Tawbi C. (1996). ADACTIF : Extension d'un SGBD à l'Activité par une Approche Procédurale Basée sur les Rendez-vous. PhD of the Paul Sabatier University - TOULOUSE III - December 1996.

Net-based Open & Distance Learning
- Experiences from the MECPOL project -
(Models for European Collaboration and Pedagogy in Open Learning)

Bodil ASK
Agder College, Norway
Bodil.Ask@hia.no

Harald HAUGEN
Stord/Haugesund College, Norway
Harald.Haugen@hsh.no

Introduction

This paper is based on the MECPOL-project which reports on 'Models for ICT-based open and distance learning'. MECPOL is concerned with collaboration between European universities in the provision of open and distance learning opportunities, especially in situations where information and communications technologies (ICT) are used. The focus of MECPOL is upon collaboratively provided ODL using ICT networks.

In addition to giving a survey of existing models for institutional co-operation regarding ODL, this project has as one of its main outcomes a *guideline* for developing and implementing an open and flexible, net-based learning environment. This paper will deal with experiences from the development of the guidelines which is based on user-trials/ field-trials, iterative evaluation as well as theoretical considerations.

The guideline

The first version of the *guideline* has already been implemented as an interactive, netbased *international course*. This course, preliminary named as 'Pedagogy in Open Learning', has become a compulsory part of the IT-curriculum for teacher training by several of the participating institutions and has been accepted for credits (3 ECTS - European Credit Transfer Systems) in different countries.

In this course experiences from developing and implementing net-based ODL are discussed and advice is given to topics like:

- Practical introduction into the use of net-based learning environments
- Theoretical approaches: Open learning, Computer supported Co-operative Learning, System Dynamics etc.
- Technological approaches
- Services on the Internet
- Design and implementation of a learning environment
- Pedagogical approach
- Hypermedia
- Design and development of Open Learning material
- Integration of ICT into curriculum
- Organisation and Economy of ODL activities
- Evaluation and quality control
- Life Long learning
- Case studies

The intention is to introduce some concepts and models relevant to ICT-based ODL, based on experiences from participating institutions and on-going work..

The guideline-course is based on survey and user trials which have been evaluated and recorded through an open, international learning and collaboration experience. Experiences from distribution of learning material as well as contacts between students and course providers, is mainly based on the use of Internet and related communication systems (WWW, Netscape etc.)

The Learning Environment

The main idea behind presenting a guideline as an interactive course are based on the *learning-by-doing* principle. The learning environment in which the course takes place is created as one example of how to design and implement a virtual learning environment using multimedia. Within this learning environment, use of hypermedia is described as well as used as tools for management of the environment. Furthermore, collaboration between teachers at different institutions occurs as an interesting component in this environment. A variety of professional profiles at the co-operating institutions strengthens the whole environment and even makes teachers become 'students'.

In a net-based learning environment like this the role of the teacher is no longer defined by tradition (if such exists nowadays?) 'Normally' the teachers have the major responsibility for what and how the students learn. But to-days learning environment opens up for a discussion of what the role of the teachers is going to be: A sage on the stage or a guide on the side? A virtual learning environment seems to indicate that teachers' tasks are moving from lecturing and 'teaching' towards supervising and assisting students. In a learning environment where courses are collaboratively developed, the new role as a guide will fit a number of teachers much better, giving them support, inspiration and job satisfaction.

Target Groups and Expected Benefits

Target Groups: Developers and providers of ODL, including teachers and students at universities and colleges, intending to work in the field of ODL, developing opportunities for professionals in need for upgrading or new knowledge and other candidates in life-long learning situations

Benefits: International exchange of knowledge and ideas. In the first user-trials a total of around 70 - 80 'learners' are registered for the international courses (in English), while larger groups are registered as participants in parallel, national courses (in Greek, Norwegian - and in English). The courses are 'not located', but developed and made available on the Internet.

International collaboration and joint development implies higher and a broader spectre of professional skills and backgroundss. Models developed as outcomes from MECPOL, are being further applied in other ODL projects, thus reaching a larger audience and making more learning *'available when and where it is needed'*. Theories elaborated and presented as part of 'pedagogical models of ODL' are now worked into the curriculum for teacher education at partner institutions.

References

Haugen, H. & Ask,B.(1997) Telecom and Open Learning - a challenge to institutional co-operation. *Telektronikk ¾.96 Tele-education. ISSN 0085-7130 p. 34 -37.*

Retalis, S. , Skordalakis, M., Haugen, H. & Ask,B. (1997) A Sage on the Stage or a Guide on the Side? Teacher's role in the University of 2000. ICTE-97 Conference, Oslo, Norway. In Press

MECPOL-project: http://idb.hist.no/mecpol

MECPOL/WP2 (1997): Models of ICT-based ODL; http://www.lancs.ac.uk/users/edres/research/mecpol/w2p2main.htm
MECPOL/WP1 (1997): Models for collaboration; http://www.lancs.ac.uk/users/edres/research/mecpol/w1p1main.htm

Team Development Of Web-Based Applications: Experiences In A Software Engineering Course

George E. Atkins
Computer Science Department
Southwestern Oklahoma State University
Weatherford, Oklahoma (USA) 73096
atkinsg@swosu.edu

Introduction

Software Engineering courses in undergraduate Computer Science education generally include a team project as a major course component. The relevance of team projects has been well-documented [Scott, et al, 94], and Web-based applications are increasingly popular [Sanderson, 97]. Team projects produce appropriate work environment experience for students. Additionally, Web-based applications provide a vehicle for undergraduates to encounter client-server architectures, to study human computer interaction, and to develop applications readily accessible to the Internet community.

This work-in-progress paper presents the status and progress of the team development of Web-based applications in a Fall, 1997, undergraduate Software Engineering Course. Students use HTML, Java applets, and CGI scripts to produce educational applications accessible via the World Wide Web. The applications involve the interactive collection, analysis, presentation, and archiving of user-provided data. The three-team structure of each project is somewhat unique with a client-side team, a server-side team, and an application definition, testing, and documentation team.

Overview

Recent Web-based applications appearing in the literature include "Programmed Instruction" [Kjell, 97] and QUIZIT [Tinoco, Fox, and Barnette, 97]. These applications involve a high degree of sophistication, but the basic concepts are appropriate for undergraduate projects. Client-side applications perform data presentation and data collection, primarily using HTML forms and Java applets, and provide the user interface. Server-side applications are developed with CGI (Common Gateway Interface) scripts to analyze and archive the data presented by the client-side application and return results to the user. The utilization of CGI scripts in Web-page development appears in much current literature [Murthy, 1997], and the combining of Java and CGI scripts, as well as the decision of using client-side versus server-side processing, is currently being described [see Pierce, 1997].

Web-based applications may be characterized as in the following figure [Fig. 1]. Form validation and data presentation processing is performed on the client-side, whereas processes requiring access to stored data or entails high performance requirements are assigned to the server-side for processing. Utilization of both Java applets and CGI scripts allow for suitable division of processing between the client and the server.

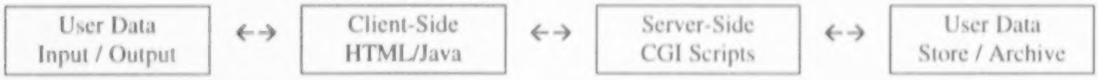


Fig. 1: - Characterization of Client-Server Interactive Web-based Applications

Projects

Two Web-based applications are developed in the Software Engineering course. The primary team project developed by the students is a "Program of Study Analyzer." This Web-based application allows a student, via the World Wide Web, to respond to questions about major and minor programs of study, courses taken and in progress, and preferences for course load. The user then receives (and interacts with) a plan of study using any web-browser. This application provides the undergraduate students a full range of client-server processing requirements for the project team. Students define and understand the application requirements, develop the client-side user interface, and develop the server-side processing of user-supplied data. The design and implementation of project testing procedures form a major part of the Software Engineering effort.

A second class application, provided time is available, is a rudimentary "on-line" tutorial and quiz system. A user responds to questions about HTML and web page development, is led through tutorial material based on responses, and receives quiz questions to measure progress. QUIZIT [Tinoco, Fox, and Barnette, 97] is utilized as an example "mature" process. Each of the applications developed by the undergraduate Software Engineering class demonstrate "intelligent" interaction and performs data collection for archive and data analysis for performance information.

The nature of the Web-based applications is quite compatible to a three team approach to project development. A client-side team develops applications to provide the user interface and to perform data presentation and collection. These applications are developed using HTML, forms, and Java applets. The more advanced computing students in the class comprise the server-side team. Their task is the development of the CGI scripts (using Perl in a VMS environment) to analyze and archive the user data presented by the client-side application and to return results to the user. Students somewhat less experienced and proficient in programming techniques comprise the essential applications definition, testing, and documentation functions. These rather small applications are developed using a prototype approach. The course progress, applications, and results are maintained and updated on a course web page as suggested by Veraart and Wright [Veraart and Wright, 96]. The course URL is http://www.swosu.edu/academic/compsci/se.htm.

References

[Kjell, 97] Kjell, B. (1997). Programmed Instruction using Web Pages, *The Journal of Computing in Small Colleges*, 12 (5), 113-123.

[Murthy, 97] Murthy, N. (1997). Using CGI Scripts to Promote Faculty-Student Communications, *The Journal of Computing in Small Colleges*, 12 (5), 124-131.

[Pierce, 97] Pierce, B. (1997). Combining Java and CGI Scripts: Communicating Between Client and Server, *Web Techniques*, 2 (9), 49-52.

[Sanderson, 97] Sanderson, D. (1997). WWW Application Development as a Software Engineering Project. *The Journal of Computing in Small Colleges*, 12 (4), 341-344.

[Scott, et al, 94] Scott, T., Bisland, R., Tichenor, L., and Cross, J. (1994) Team Dynamics in Student Projects, *SIGSCE Bulletin*, 26 (1), 111-115.

[Tinoco, Fox, and Barnette, 97] Tinoco, L., Fox, E., and Barnette, D. (1997) Online Evaluation in WWW-based Courseware, *SIGCSE Bulletin*, 29 (1), 194-198.

[Veraart and Wright, 96] Supporting Software Engineering Education with a Local Web Site, *SIGSCE Bulletin*, 28 (1), 275-279.

Designing Hybrid CD-ROM/Web: A Reality Check

William G. Barek
Director of CITD
CITD - Centre for Instructional Technology Development
Bladen Library
Scarborough Campus
University of Toronto
1265 Military Trail
Scarborough, Ontario, Canada, M1C 1A4
Tel: (416) 287-7505 Fax: 416) 287-7507
e-mail: barek@scar.utoronto.ca
WWW: http://citd.scar.utoronto.ca

Introduction

When discussing multimedia development and delivery we come across a heated debate between two different camps. Those promoting and defending CD ROM based technologies and others promoting exclusive Web based technologies. This paper explores some issues surrounding these debates and it looks at the Hybrid CD ROM/Web technology that takes the best from both. When considering the use of multimedia technologies, issues concerning authoring, delivery and content updates are the most important. In particular, questions focus on whether the authoring tools are powerful and easy to use with both formats and whether there is a real benefit from using this technology.

Why use hybrid CD-ROM/Web technology?

CD-ROM has been considered by many as a good and inexpensive vehicle for delivering interactive multimedia content. The major criticism of CD-ROM is that its content is static and quickly becomes out of date. On the other hand Internet and in particular the World Wide Web, offers incredible flexibility, real time information and distributed collaboration. On the negative side, the bandwidth and authoring capabilities are very often criticized and are often seen as the major obstacles in the creation and delivery of multimedia content.

It has often been suggested (Ozer 1997) that improvement in the Internet bandwidth will eventually kill the CD-ROM. It is also widely agreed that wide adoption of Web in the developers community will improve the authoring tools as well. If this will eventually happen and if so when is open to speculation. Meanwhile, the advantages of utilizing both technologies to create hybrid CD-ROM/Web content should be considered as a bridged solution.

Different types of Hybrid CD ROM/Web Designs

It has been estimated (Cole 1997) that in the mid-1996 there were 350 or more hybrid titles and that by the end of 1997 there will be around 3500 tiles available. Many are from well established companies such as Microsoft, Voayager, Grolier, Dorling Kinderslay and others. Careful evaluation of these products will allow the seperation of marketing hype from the useful tools that will improve the development and delivery of multimedia content.

The major types of hybrid CD-ROM/Web designs to consider are:

1. CD-ROM media content (video, audio, large graphics) accessed directly from web browsers (could be either local or remote mode).
2. Interactive multimedia titles with simple link to a web site by launching a web browser.

3. Interactive multimedia titles with links to the various media content update site or sites (e.g. Microsoft Baseball, Cinemania 97).
4. HTML-based CD-ROM with HTML structured content (e.g. Encyclopedia Britanica 2.0).

Each format has its own strengths and weaknesses and we can choose those elements that are particularly suitable to our application.

Creating hybrid CD-ROM/Web Content

There are several authoring tools available on the market today. Asymetrixis ToolBook II, Allen Communicationis QuestNet+, MarketScapeis WebCD, Macromediais Director 6.0 and Authorware 4.0 and others embrace the Internet integration more effectively and with variou ease-of use. We have chosen Director 6.0 and Authorware 4.0 for our projects because of our familiarity with the software, its suitability to our particular content design and its cross-platform capability. In particular, Director 6.0 offers good integration with Java, so that movies can be embedded as applets. Similarly, Java applets can be played within Shockwave movies. Other enhancements to the Director 6.0 include support for Active X controls, QuickTime VR, QuickDraw 3D, DirectSound, JavaScript and LiveConnect. These capabilities offer a number of advantages over other authoring tools.

Conclusion

With the advances in Internet-based technologies, such as VRLM 2.0, proposed HTML 4.0 and XML markup language, improved Java JDK and just in time (JIT) compilers, ActiveX and Java Beans the integration of CD-ROM looks very promising (Gustavson 1997).

At the CITD, we have started to create a *University of Toronto at Scarborough Promotional CD ROM*. It will consist of promotional video clips, animation and voice-over with many images - media that CD ROM handles reasonably well. On the other hand, fast changing textual information is well implemented on the Web. Placing media intensive part on the CD ROM (such as video and animation) and linking the changing textual information (such as course calendar, timetables and other information) presently available on the Web site to it seems to be the best solution for now. The projected completion of this Hybrid CD ROM/Web project is July 1998 and complete report on this project will be generated as well. At the same time we are also evaluating the possibilities of using this technology in several distance education projects. Similarly, many publishers today consider CD-ROM not as exclusively standalone product, but rather see it along with the Internet, as a hybrid strategy in their publishing effort (Cole 1997).

References

Cole, J.O.
Publishersis Strategies for CD-ROM/Internet Publishing
EMedia Professional: Vol 10, No.1, January 1997

Gustavson, R.
Standards from the Web and the CD-ROM Time Machine
Emedia Professional: Vol.10, No.3, March 1997.

Ozer, J.
Talkingi 2000: Will Breakthrough Internet Bandwith Sink CD-ROM?
Emedia Professional: Vol.10, No.5, May 1997.

Jeremy, C.
CD-ROM, Web hybrid Speeds Up Multimedia
Web Week: Vol.2, Issue 13, September 9,1996

"Hangar Flying" as story-based instruction: Capturing expertise via online video libraries

Benjamin Bell, Sholom Gold, Danielle Kaplan

Institute for Learning Technologies, Teachers College, Columbia University

Many communities of practice rely on first-person narratives as a means of conveying knowledge from the experienced practitioner to the novice. These "war stories" serve a valuable instructional purpose, enhancing basic concepts with richly contextual accounts. We report on a story-based instructional system that provides for an engaging, exploratory interaction between experts and learner, and introduce a tool for wide-area dissemination and collection.

First-person narrative is a traditional and widely accepted means of instruction. Because stories and anecdotes contain basic information that is linked in a natural and contextualized way, they serve as discrete "packets" of information that can be readily shared by experienced mentors. In many professional contexts, storied instruction becomes an intrinsic element in the enculturation and education of novice practitioners.

Stories hold promise for instructional applications for several reasons. Stories encapsulate ideas and concepts within a rich set of contextual cues that can render the ideas and concepts immediately accessible to the learner. Stories go beyond expressing a skill and provide for new information to be integrated into a set of conditions under which that skill is appropriate (or inappropriate) to apply. Abstracted and decontextualized knowledge is often difficult to master because the student is burdened with inventing situations in order to grasp the abstraction. Stories, on the other hand, are a real-life instantiation of some abstract set of principles. Reading a decontextualized explanation of a phenomenon is not nearly as convincing (or memorable) as hearing a credible colleague or mentor discuss first-hand knowledge of that phenomenon. Moreover, because human memory is more adept at logging away specific episodes than at recalling facts and axioms, stories are more readily retained precisely because they associate a set of concepts with a corresponding sequence of events. Because they are memorable and often provocative, stories also provide motivation to the learner.

The utility and appeal of such first-person accounts has led some researchers to develop technological approaches that introduce a story-telling factor into the teaching equation. We are exploring the instructional benefits of stories when coupled with the reach of the web (augmented with streamed media and a dynamic database) in a prototype application titled "The Aviator's Story Archive" (edtech.tc.columbia.edu/comet/crew.html). ASA is a web-based resource for pilots who wish to maintain proficiency in their grasp of various atmospheric phenomena that pose potential flight hazards (e.g., turbulence, icing). The information is available in the form of "war stories" that pilots share about personal encounters with these conditions (the sharing of such stories among pilots is often referred to as "hangar flying"). The system is available as a web site and consists of four main components: a Story Logbook, a Dispatch Room, a Conference Room and a Crew Lounge.

The Story Logbook is an archive of video stories in which professional pilots recount first-hand experiences about flying in various kinds of weather conditions. The learner is able to access the database through a series of questions and responses. That is, each question is linked to a video answer and several follow-up questions. By following these thematically linked connections among videos, or by viewing a master list of all questions that may be posed, visitors can follow individual paths through the archive according to personal preferences. The Conference Room, currently underdevelopment, introduces the important dimension of participation among the community by offering a channel through which people can contribute their own experiences to the archive. The Dispatch Room and Crew Lounge, also under development, will serve as sources (and links to sources) of live weather information and weather-related instructional materials, and as an area for chat rooms and discussion groups, respectively.

Bartering and Gaming for Education on the Web with Non-convertable Virtual Currency

by Prof. Dr. Gary McI. Boyd
Education Dept. Concordia University, Quebec, Canada.
<Http://alcor.concordia.ca/~boydg/>

Introduction

The Internet and the Web are good media for education, but not as good as they can be. Cyberspace allows representations of people and knowledge from all over planet Earth to come together interactively. This sometimes is very educational - when the people involved learn to overcome old learning habits and co-construct better understandings of one-another's worlds. This does not happen as often as it should because people are unable to quickly and easily find just those other people and those tools and that particular information which they need to learn what is immediately important for their own, and for our global society's development.

One obvious strategy being pursued is for educational institutions to market their conventional courses on the net. This provides some quality control, and some economic sustenance. The great disadvantages of this old-wine in new bottles strategy are: 1) What is offered is not JIT/JOT, not Just In Time, Just On Topic learning. The course packages are too big, and slow, and include too much irrelevant stuff for most people most of the time. 2) The services are too expensive if they give really good personal learning tutorials on line. The rich benefit still more than the poor. 3) If the packages are just cheap "canned" presentations, they may be affordable enough, but are too difficult for many people to understand properly.

The other obvious strategy which has been employed since ARPANET days is merely to carry on by "free give and take". Put up whatever you have to offer, and let anybody make what use they can of it. Also make appeals for help JIT/JOT, and generously respond to such appeals. This has worked very well in the scientific and some educational communities where people are socialized to respect one another's work and authorship, and where the communities involved are small enough so that they can operate as "Grace & Grudge" networks. That is, everything is freely given until it is found that somebody "rips-off" other's work without acknowledgment or return contributions. Then the community can hold a grudge against the opportunist, at least until behaviour improves. On the open WWWeb these strategies don't work well enough. People abuse the commons of the Web by "pushcasting" huge multi-media files; noise and crimepetitive activities abound. The final commercial solution is to introduce micro-payments systems - so that everything is paid for bit by bit, but this favours the rich, and does not promote synergetic learning communities.

Towards a Better Medium of Exchange and Approbation

The commercial "solutions" are not so good for education or science because most of the people who really need to learn are too economically deprived ("free-markets" always tend to make the poor poorer and the rich richer through positive feedback) to be able to pay in "real" money.

On the other hand most learners and many teachers have their own good-quality attention-time "QuattT" (ultimately the only thing of value to human beings) to barter. With the web and public key encryption certificates[Godin, 1995] it has now becomes possible to exchange QuattT credits directly without government or corporate currencies. This is very important because closed community currencies promote synergism and mutual support and re-investment within the community, whereas open currenciews tend to drain off resources[Lietaer, 1997]. Simple barter or even a market system based on QuattT is not good enough because of the noise in the web this new tower of Babel.

Go-betweens are needed to help make optimal connections among: tutors and learners and tools. What sort of go-between systems are necessary? possible? desirable? My current[Boyd. 1997] silver-bullet scheme: is for a Web-based lifelong-learning money-free brokerage for high-quality relevant attention-time barter, to be mediated by public-key-encrypted we-owe-you ephmeral credit vouchers. The tricky bit is how to arrange it to promote `wise-collective-being' not just more of the now rampant opportunistic crimepetitive individualism.

Barter with "QuattT Credits"

One possibility is for people -librarians, human cybermediators etc. to come to one's aid - for payment in QuattT. This can meet many if not most persons' needs in a big enough network, since different people have widely different areas of expertise -all of which someone somewhere probably needs. Promises of aid-in-kind in cyberspace need to be backed up by authentication certificates from some trusted institution because there is no real place where participants are bound to ongoing contact with one-another. The authenticating institution has to be trusted by both requestors and providers to keep honest records of real identities, and of who did what for whom and how well promises were kept, and how many promises each has made, and to maintain privacy except in agreed respects. The proposal here is that modest (40 digit say, since this is internationally exportable) RSA public Key encryption be employed to: a) Verify and record in trusted educational institution registries, who is doing what with what &/or with whom, for how long. b)Send estimated quality appraisal acknowledgment credits from anyone who benefits to a public registry and to whoever supplied the benefit(Of course mutual benefit may be involved in a learning conversation so that both participants may send QuattT acknowledgments to each other). c)Enable payment for bandwidth/time using QuattT credits. d) enable public reputations for specific kinds of helpful contributions to be exhibited based on the histories of QuattT "earnings". e) Keep track of expiry dates of QuattT credits, and of each participant's history of fulfilling sho's commitments by redeeming such credits.

Relevant Credibility Status Knowledge Development Gaming

Scientific research proceeds as much by people presenting papers with conjectures, or refutations of others conjecture), at scientific conferences and obtaining the attention of high status scientists,. If they do this their status goes up, and publication and grant opportunities open up. This same credibility status/attention-time gaming mechanism can be implemented on the Web as a normal way for undergraduate and adult learners to develop their lnowledge and status with peers in their specialty. But this can be done only if the attention-time (QuattT) is recorded and the accumulations are publicly displayed. This can easily be done with appropriate add-ons to browsers, and with identification certificates from some educational institution [Boyd, 1993].

What Next

All this is certainly technically possible with public-key encryption, and appropriate accountability institutions. Whether such cybermediated Just-In-Time/Just-On-Topic tutoring paid for by QuattT Credits is a politically economically and paedagogically practicable form of Web-based community education remains to be discovered through prototype experiments such as the one I am currently conducting with SAVIE(Societe' pour l'apprentissage a vie)[SAVIE, 1997]. A longer paper on this topic is available from the "publications" section of my homepage[Boyd, 1997].

References

[Boyd. 1993] Boyd, G.M. (1993). Reflections on the Future of Distance Education by C. Dede. American Journal of Distance Education, 10 (2) , 42-47.

[Boyd, 1997] Boyd, G.M. (1997) Distributing Real Learning Conversations With Virtual Money. <http://alcor.concordia.ca/~boydg/ drboyd.html#rec>.

[Godin, 1995] Godin S.(1995) Presenting Digital Cash. Indianapolis Ind. Sams.

[Lietaer, 1997] Lietaer, B.A.(1997) <www.transaction.net/web/conf/confintro>.

[SAVIE, 1997] SAVIE (1997) , <http://WWW.SAVIE.qc.ca>.

An Interactive, Interdisciplinary Web Site Template for Elementary Education Using Remotely Sensed Imagery

Ginger Butcher, SGT, Inc., USA, ginger@ltpmail.gsfc. nasa.gov
Josephine To, Science Systems and Applications, Inc., USA., joto@ltpmail.gsfc.nasa.gov
Jill Twetten, SGT, Inc., USA, jtwetten@pop900.gsfc.nasa.gov
Jon Avila, Mary Washington College, USA., javila@gsfc.gsfc.nasa.gov
Houri Tamizifar, SGT, Inc., USA., htamizif@pop900.gsfc.nasa.gov
Karen Kaye, USGS Biological Resources Division, USA, kkaye@nbs.gov
William Webster, NASA Goddard Space Flight Center, Code 922, USA, webster@ltpmail.gsfc.nasa.gov

Introduction

A web-based template was created during the process of developing an interactive web site for teaching elementary school students about remote sensing and biodiversity. Our requirements are twofold: a highly interactive, challenging environment, and a high quality educational resource usable in the classroom. First, to create an interactive environment, we developed a navigable interface consisting of geographically contiguous, remotely sensed images. Remotely sensed images present an ideal foundation because they can refer to an extensive range of content areas. A story narrative is used to challenge the student to explore the interdisciplinary content. Java script adds interactivity to the interface and feedback to the student. Secondly, most multimedia educational programs overlook practical application in the classroom. The site provides resources for reinforcing and integrating educational content. These resources are indexed by educational outcomes which aid teachers in customizing lesson to accommodate their individual student and curriculum requirements. The site forms a viable educational template to present interdisciplinary content through the combined use of remote sensed imagery, story narrative, interactive environment, and teacher resources.

Background

The educational objective is to show an application of remote sensing through NatureMapping to demonstrate how scientific discovery from space can enhance life on earth. Earlier educational prototypes were linear, offering little interactivity to effectively engage elementary students in educational content (Masuoka, 1996). These students need a challenging environment that is non-linear and highly interactive and comparable to video games on the market today. Also missing from many edutainment multimedia titles in the practical application of educational content in the classroom. The site must introduce students to educational content within an interactive site and provide teachers with resources to reinforce these concepts in the classroom. To meet these objectives, a template was created.

Creation of a Web-based Adventure Environment

The story introduction sets the stage for the interactive adventure. This linear and interactive story develops the characters and challenges students to accomplish a mission. The mission prepares the student for the interactive adventure portion of the site.

The interactive adventure environment uses three tiers of web pages. A tier is a coordinate system comprised of a Landsat mosaic image cut into squares. A web page created for each square. The first tier introduces the student to the remote sensed image and the educational content of that square, but does not allow navigation to any of the adjacent squares. The combination of story narrative and Java scripting encourages the student to click on the image to investigate. Upon investigation, the second tier in the same geographic location is displayed. The educational content is presented on this second tier. The student continues to the third and final tier where they receive a directional clue. The student may then navigate north, south, east or west to the next square.

The clues direct the student to a series of lessons. The programming of "cookies" into the web pages allows the browser to store information about the students' progress. A "cookie" is variable programmed into the web page for each lesson. As the student visits a location where a lesson exists, the status of that cookie changes. The status of these cookies prompt the site to which clue to display. The student must visit all lessons to finish, but not necessarily visit all squares of the geographic region before they finish. Additional tools are provided for the student to help in their travels, including maps and a help button which displays a hint. When the student completes a lesson, the achievement is acknowledged and the clue for the next lesson location is revealed. This system creates a linear sequence of lessons within a non-linear interactive environment.

Educational Content and Lesson Modules

The three tiered architecture of web pages form a template for the inclusion of educational content. As the student travels through the system of squares, the geographic location and remotely sensed imagery will reference educational content. Remote sensing is a scientific tool which has the ability to reference a wide variety of content areas. Examples of disciplines that utilize remote sensing data of the earth's surface include geology, biology, urban planning, and hydrology. Lesson modules are included to highlight the critical concepts pertaining to the overall educational objectives of the site. These lesson modules are comprised of web animations, illustrations and lesson narrative. The supporting story narrative, in conjunction with the geographic location, forms a foundation to integrate a variety of educational content.

Teacher Resources

The teacher resources consist of thematic units which facilitate the integration of site content into the classroom. Within each unit, lesson plans, lesson modules and content pages are organized by national science educational outcomes. This enables teachers to construct lessons to accommodate their individual student and curriculum requirements. The lesson plans provide teachers with ideas on how to reinforce the content and provide practical applications of the complex concepts covered in the interactive portion of the site.

In addition to lesson plans, direct links to the lesson modules and the content pages are available. These links give teachers direct access to the lesson module animations to assist in their classroom presentations of such complex concepts. Teachers can also access the content map which describes the content for each square the students may encounter. This flexibility allows teachers to use the animations to supplement lessons in their classrooms.

Future Enhancements

This project is scheduled for release June 1998. Enhancements to the program will be discussed after the completion of beta testing in May of 1998. Such enhancements include the randomization of the location of the lesson modules within the coordinate system. This would allow students to replay the adventure and view content not visited on their first trip.

Reference

Masuoka, P.M., Samadi, S., Khosand, M., and Foresman, T., 1996. Developing a Mulitmedia Package for Remote Sensing Education. 1996 ASPRS/ACSM Annual Convention and Exposition Technical Papers, Vol.I pp. 531-538.

An Aglet-based Distributed Web Indexer

Marco Chirico, Fabrizio Giudici, Andrea Sappia and Anna Marina Scapolla
Department of Biophysical and Electronic Engineering
University of Genoa - Networking Competence Center
Via All'Opera Pia 11/A - 16145 Genoa, Italy
Phone/Fax: +39-10-3532192/2175
E-mail: sappia@dibe.unige.it

1. Why Mobile Agents to perform Web searches?

The World Wide Web (WWW) [WWW Cons] has definitively become a standard way for distributing documents of different kind. This explosion of information leads to the need for document indexing and retrieval tools like search engines (Lycos , Yahoo , Altavista, Infoseek , etc.).

A WWW search engine is mainly composed of a *robot* that explores the web sites travelling across HTML hyperlinks and an *indexer* that extracts the information from the documents building tables of their contents.

In the current centralised model the program, performing both the search and the indexing, is executed on one or more hosts that appear to be a localised spot of Internet. When a robot tries to index a document from a site, first it downloads the document, occupying network bandwidth, then builds the indexing information, using local computational power. Today's trend is to have every kind of objects (like sounds, animation, database tables) in a WWW page, thus keyword search will rapidly become inadequate and more powerful techniques will be required.

Distributed approaches seem to be promising solutions to the need of more powerful and customised Web searches. Many distributed computational models have been proposed, but they lack a general interaction and communication model (e.g. Java [Gosling & McGilton]) or the capabilities to transfer code and the right distribution granularity (e.g. the "interacting distributed objects" of CORBA [CORBA]).

The Mobile Agent model provides both mobile code and interacting objects. A Mobile Agent [Harrison et al.] is an object that can migrate from host to host executing a sequence of operations without any supervision, and it is characterised by a *data state*, a *code state* and an *execution state*.

IBM is working on combining Java *applet* [Java Tutorial] technology with agents to generate "*aglets*" [Lange & Chang]. IBM is championing this as a standard for implementing Mobile Agents in Java and has already submitted a proposal to the Object Management Group.

2. The Model

This paper presents a model, named *Bees*, that foresees WWW servers co-operating with the search engine since they host and execute both the robot and the indexer programs in form of a Mobile Agent (figure 1). We name the originating host that commits the search "*beehive*" and the remotely executed agent "*bee*". The beehive uploads Bees on the WWW servers, then the Bees explore the sites *from inside*, and send back results to the beehive. Bees can spread over the Web returning to the beehive only when all their tasks have been completed.

Bees software is composed of an aglet (*Queen Bee*), located on the originating host, that creates and dispatches indexer aglets (*Worker Bees*) to the Web Servers. Worker Bees consist of two parts: the *Web Robot* that explores a site starting from a given initial URL and the *Web Indexing Engine* that processes WWW documents and builds the table of contents. When Worker Bees return to the beehive they store search results into the *Search Database*.

Today's robots can be unfriendly: they can explore lots of documents very quickly, producing the so called "rapid fire", consuming system resources and network bandwidth. With the *Bees* model bandwidth is saved since table of contents are usually smaller than WWW documents and the footprint of the Worker Bees is usually small, since WWW support is already present in standard Java libraries. The WWW servers can control the way Worker Bees operate inside them, giving limited amounts of CPU power or scheduling them for execution when the load is low (e.g. overnight).

Bees are implemented with the Java language, thus they are fully portable among different platforms, and their indexing capabilities can be easily customised: during its activity the Worker Bee builds a complete representation of the hosting Web site, that can be easily accessed from the Indexer Engine. In this way, complex searches exploiting relationships among different pages can be implemented. Sophisticated Bees-based Search Engines can implement customised searches by specialising the Indexer Engine and complex data structures resulting from searches can be easily carried back without the need of dedicated protocols, thanks to the automatic serialisation provided by the Java/Aglet runtime.

The *Bees* model is secure: Worker Bees cannot *sting* since the Java/Aglets Security Manager prevents them from executing malicious code on the visited hosts; moreover access restrictions policies can be implemented.

Summing up, this model exhibits a number of good properties when compared with the standard centralised approach: bandwidth saving, computational power distribution, system resources saving, batch scheduling, access control policies.

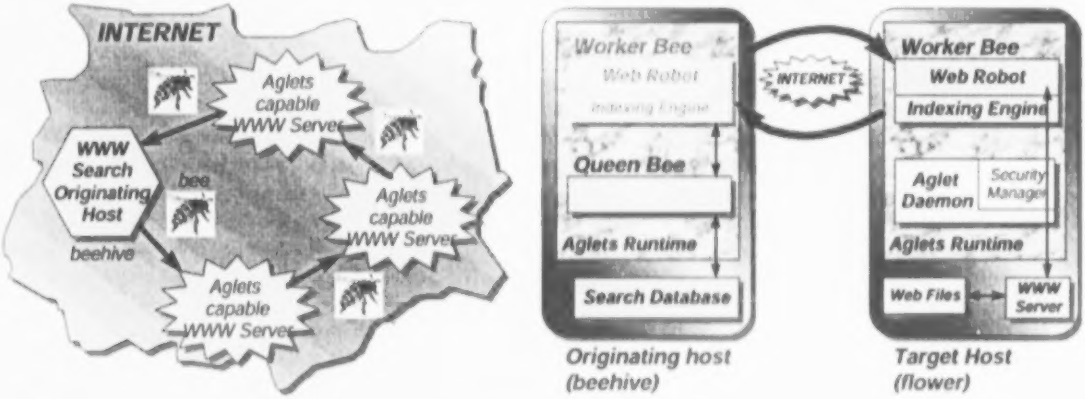

INTERNET
Aglets capable WWW Server
WWW Search Originating Host
bee
beehive
Aglets capable WWW Server
Aglets capable WWW Server
Worker Bee
Web Robot
Indexing Engine
Queen Bee
Aglets Runtime
Search Database
INTERNET
Worker Bee
Web Robot
Indexing Engine
Aglet Daemon
Security Manager
Aglets Runtime
Web Files
WWW Server
Originating host (beehive)
Target Host (flower)

Figure 1 - Bees Model and Architecture

3. Future Directions and Conclusions

Presently the *Bees* prototype connects with HTTP servers by opening local sockets. It can be optimised by directly interfacing with the HTTP server, saving operating system resources on the visited hosts. A proper API should be designed for this purpose, and Jigsaw [Baird-Smith], the freely available HTTP server by the W3 Consortium, appears to be the most suitable choice, since it is entirely written in Java.

An authentication model for Worker Bees based on cryptographic techniques will be implemented when security Java/Aglets support will become more stable.

Bees will be soon available in Public Domain at their home site *http://www.ncc.dibe.unige.it/Bees*. While their profitable use through the Internet will have to wait the standardisation of the Aglet proposal and the diffusion of the supporting software, we believe that they can be soon used within Intranets.

References

[WWW Cons.] The World Wide Web Consortium, *http://www.w3.org*

[Gosling & McGilton] J. Gosling, H. McGilton, "*The Java Language Environment - A White Paper*", Sun Microsystems Inc. *http://www.javasoft.com/doc/language_environment*

[CORBA] Advanced Computing Laboratory, Los Alamos National Laboratory, "*CORBA and OMG Information Resources*", *http://www.acl.lanl.gov/CORBA*

[Harrison et al.] C. G. Harrison, D. M. Chess, Aaron Kershenbaum, "*Mobile Agents: Are they a good idea?*", IBM Research Division, *http://www.research.ibm.com/massive/mobag.ps*

[Java Tutorial] "*The Java Tutorial*", Sun Microsys. Inc., *http://www.javasoft.com/books/Series/Tutorial/index.html*

[Lange & Chang] D. B. Lange, D. T. Chang, "*IBM Aglets Workbench - Programming Mobile Agents in Java*", IBM Corporation, *http://www.trl.ibm.co.jp/aglets*

[Baird-Smith] A. Baird-Smith, "*Jigsaw Overview*", The World Wide Web Consortium, *http://www.w3.org/pub/WWW/Jigsaw*

BEST COPY AVAILABLE

A Monitor System for Cisco Router Interfaces:

A Case Study for a University Metropolitan Area Network

Enrico Commis
Dipartimento di Matematica, Università di Catania,
Viale Andrea Doria 6, I-95125 Catania, Italy
E-mail enrico@dipmat.unict.it

Luca Gentili
CINECA,
Via Magnanelli 6/3, I-40033 Casalecchio di Reno (BO), Italy
E-mail lucag@cineca.it

Introduction

The fast growth in network links and Cisco routers traffic at the University of Catania MAN has led to the necessity for fast and efficient retrieval of traffic informations from the routers. In addition, since the University of Catania MAN has only one 768Kbit/s line to the outside world, it is very important to evaluate the link performance.

In this paper we describe a *Software Prototype for a Monitor System* (SPMS) of the traffic load on network-links. All the information is accessed and visualized inside of a HTML compliant browser: SPMS offers graphic representation on the spot (GIF images) of traffic of the monitored network connection, embedded into webpages which can be viewed from any web-browser.

In addition to a daily view, SPMS is able to create visual representation of the traffic seen during the last 24 months as well as generate a monthly traffic summary. This is possible because the tool keeps log of the relevant data for all the traffic seen over the last two years.

The system integrates: a Perl5 script which uses SNMP to read and log the traffic counters of our Cisco routers, with several C programs to elaborate on the fly GIF images representing the traffic on the monitored network connection.

SPMS architecture

Following the OSI guidelines, SPMS uses the *Simple Network Management Protocol* (SNMP) to read traffic counters of the Cisco routers of the University of Catania MAN.
To retrieve traffic information from a router interface, SPMS asks the SNMP agent of the monitored router (via a SNMP *Get* operation/message) for an instance of several MIB-II objects: *ifNumber, sysUpTime, ifDescr, ifInOctets, ifOutOctets.*
To allow the SPMS to retrieve an object instance from the router SNMP agent, a Perl5 script process (*AppDaemon*) is executed at scheduled time (generally every 5 or 10 minutes) on the web-server workstation (currently a Sun Netra workstation running SunOS 5.5.1 operating system).
As input, *AppDaemon* requires an ASCII file containing relevant information of the routers to be managed.
AppDaemon stores traffic load of routers' interfaces into log files. This log is automatically consolidated, so that it does not grow over time, but still contains all the relevant data for all the traffic seen over the last two years.

Deploying SPMS on the Web

Most W3 servers provide one or more APIs for integrating new and existing applications. The most well known of these APIs is the *Common Gateway Interface* (CGI), and although some servers (such as Netsite for example) additionally provide a specialized API, CGI is currently standardised and supported by all major W3 servers.
We now describe our experiences of using CGI API to deploy visual representations of traffic load on network-links on the Web.
The *application front-end* is simply a Web browser and a CGI Perl5 script (*front-end* script) that presents a FORM document to select (via pop-up menus) a graphical view of a monitored day.
The user query is a conjunction of two attributes:

- **Link**: Logical name of the network link (router interface);
- **Date**: *Day, Month* and *Year.* In addition to the today's date given by default, the user can freely choose a day of the last 24 months.

When the form generated by the *front-end* script is submitted for processing, the browser invokes a new CGI Perl5 script (*show* script) in order to create a visual representation of the traffic load monitored during the selected day.
The *show* script first performs a filtering process to retrieve, into a temporary file, the relevant daily information from the stored log files. Then the script gives an HTML page document back to the client. This HTML page contains an IMG element which has the following format:

```
<IMG SRC="/cgi-bin/plot">
```

where *plot* is a very fast C program which generates on the fly a GIF image representing the daily view of the traffic on the selected network link (router interface). Using this approach the graph is directly embedded into webpage without any additional log data.

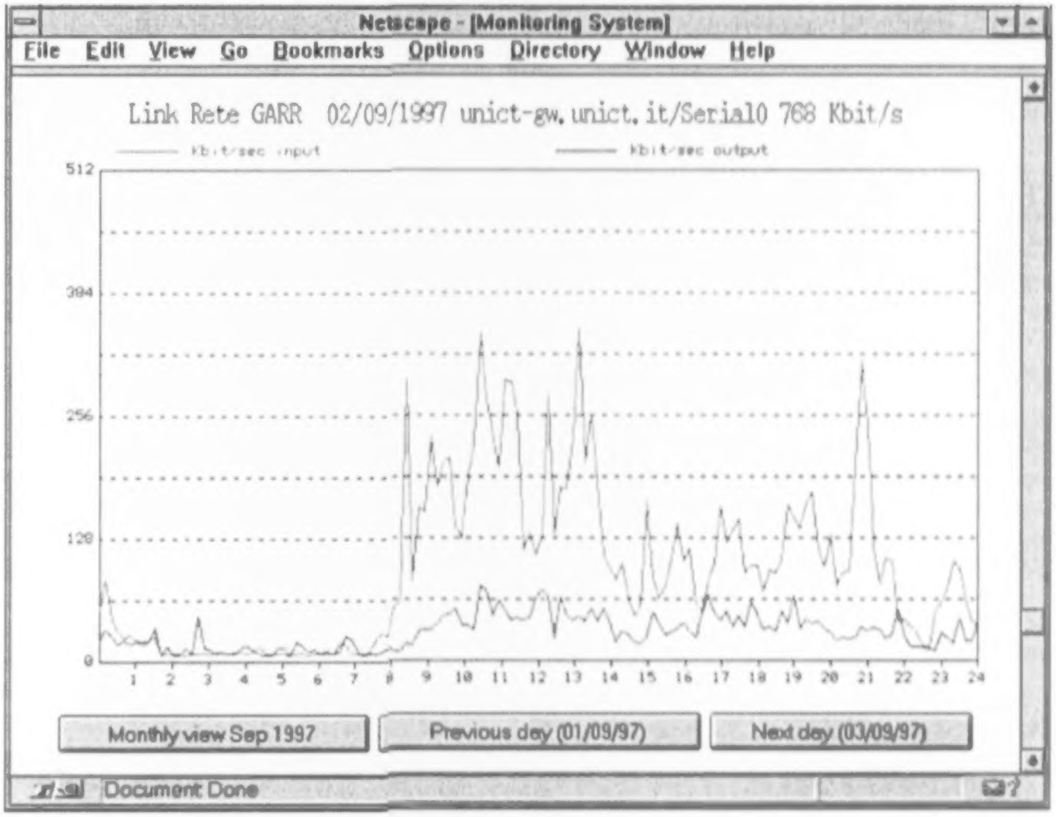

Netscape - [Monitoring System]
File Edit View Go Bookmarks Options Directory Window Help
Link Rete GARR 02/09/1997 unict-gw.unict.it/Serial0 768 Kbit/s
Kbit/sec input
Kbit/sec output
512
384
256
128
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Monthly view Sep 1997
Previous day (01/09/97)
Next day (03/09/97)
Document Done

Figure 3.1: Daily view of traffic load.

The HTML page presented by the *show* script also contains three FORM elements (push buttons) which let the user to re-submit a new query for the previous or next day (if applicable), as well as to submit a query for a monthly view of the traffic load of the specified month.

Conclusions and future work

In this paper we described the SPMS software: *Software Prototype for a Monitor System.* The system has been used in practice to monitor the traffic load on network-links at the University of Catania MAN; our experience proves that the tool is effective and useful for the management of complex networks.
Future work on the SPMS regards: monitoring of a larger number of traffic parameters; integration with tools for dynamical bandwidth allocation.

References

[Stallings 1996] Stallings, W. (1996). *SNMP, SNMPv2, and RMON: practical network management.* Addison-Wesley Publishing Company Inc.
[Wall and Schwartz 1992] Wall, L., & Schwartz, R.L. (1992). *Programming perl.* O'Reilly & Associates Inc.

BEST COPY AVAILABLE

Computational Agent to Integrate An Expert System for Design and the Internet

Jonny Carlos da Silva
Mechanical Engineering Department, Federal University of Santa Catarina,
Florianopólis-SC-Brazil- CEP 88040-900. Email: jonny@emc.ufsc.br

David Dawson
Engineering Design Centre, Lancaster University,
Lancaster- LA1 4YR- United Kingdom. Email: d.dawson@lancs.ac.uk

Introduction

The present work corresponds to a module of the Schemebuilder project, which is a software tool under development at Engineering Design Centre (EDC)-Lancaster University-UK aimed at getting better Conceptual Designs faster. In this Concurrent Engineering scenario, individuals from different areas in the industrial environment (such as, design, manufacturing, suppliers, quality control, etc.) interact during the product development from the early stages, Conceptual and Preliminary Design. Those stages are characterised by two important aspects. Firstly, as stated by practitioners of concurrent or simultaneous engineering, the decisions taken during these early stages have the greatest impact on the product life-cycle. Secondly, these stages have the highest level of information abstraction, since the design evolves from the user needs and requirements to the system specification. Therefore, the main purpose of Concurrent Engineering is to shorten the product development time, including design, keeping a better quality and avoiding rework in the later stages. The above mentioned aspects support the application of computer-based tools, such as expert system, to facilitate the interaction among the individuals participating in a product design environment.

Expert System Choice

We decided to implement this AI system through an expert system approach based on the following aspects: Development of rapid prototype (due to the time constraint of the project); Capacity to provide an explanation facility; Availability of a reliable implementation tool; and Capacity to represent symbolic manipulation. Thus, we have used CLIPS (C Language Integrated Production System, shell tool developed by NASA) as the implementation tool [Giarrantano,94]. Having defined this point, it was clear the great complexity of a design task cannot be modelled using only the Rule-Based paradigm (If A Then B), therefore a decision was made to use the COOL module (CLIPS Object Oriented Language) which allows the application of fundamental properties, such as inheritance, abstraction and assembly relationships [Silva,97a]. Although the decision to use the expert system paradigm was made much earlier, a recent survey done through a set of questionnaires posted to WEB newsgroups has proved to us that the choice is applicable and appropriate for the defined domain. The knowledge acquisition process in this project is described in [Silva,97c].

Computational Agent

The most common application for agents is gathering data to build indexes for search engines in Internet applications. Other names for these "Resource Directory Agents" include so-called robots, spiders, and wanderers. Every time a web search engine is used, the search process is carried on through files created by agents [Personal Agents,96]. There are many definitions of "agent" in the area of computer supported systems, some of them are presented as follows: *An agent is always at least postulated for every action. ...An agent is a representation which produces a change in representations in a model* [Turchin,92].

In the present work, the Expert System is composed of a set of computational agents that are tailored for specific tasks, such as: output data format for linking with simulation packages; hydraulic system troubleshooting solver; and data format in HTML. This last module is the objective of this paper. A general structure of this Expert System is presented in [Silva,97b].

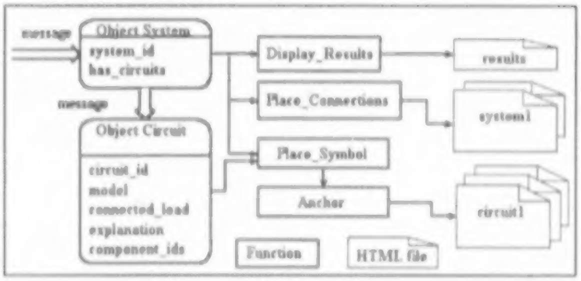


Figure1- HTML Agent Structure.

As depicted in figure 1, the agent is composed of several messages that are passed to System and Circuit Objects in the hydraulics domain allowing the generation of a set of HTML files with textual and graphical information linking the alternative hydraulic systems to their circuits. This agent is the main interface to the user and also allows the user, through hypertext facilities, to navigate among the several files that are automatically created as well as help on-line options. The expert system which embraces this agent has been intensively tested by a design consultant in USA.

Internet- Benefits and Potentials

The World Wide Web with its HTML language has quickly become a standard means for hypertext document delivery [Tanskanen,97]. Because of numerous advantages, Web tools are soon expected to be found on each engineer's desktop. Even today, some speak of 85% of the workplace computers that are already connected to the Internet using some of the Web services. The spreading of Web technology had also the interesting side-effect that many programs without a user-interface have emerged, which rely wholly on a web browser for interaction. Web tools run on all major platforms and have a high degree of compatibility with all kinds of applications. Web technology is widely known and standardised and offers good communication performance. The resulting tools have already gained the acceptance even of engineers that belong to the "late adopters" of computing and network tools [Drisis,97].

Conclusion and Future Issues

So far the expert system approach adopted in this project, combining with the massive use of the Internet for knowledge harvesting and also as an interface standard has been demonstrated successfully. At present, only the expert system output is formatted to the Web. However, plans have been made to develop a complete Internet application system. A description of the type of output generated by the agent can be found at EDC Homepage (http://www.comp.lancs.ac.uk/edc/).

Acknowledgements

This project is related to the thesis of the first author, who is also lecturer at Mechanical Engineering Department of UFSC. It is being carried out as a part of a doctorate research programme sponsored by the Brazilian Research Agency CAPES. This project is being developed at Engineering Design Centre, Lancaster University, UK, through an exchange programme with the Federal University of Santa Catarina (UFSC), Brazil.

References

[Giarrantano,94]- Giarrantano, J. and Riley, G., Expert Systems- Principles and Programming, Second Edition, PWS Publishing Company, 1994.

[Drisis,97]- Drisis, Leonidas. Wide Area Design Teams- Tools for Collaboration. ICED97, International Conference on Engineering Design, Tampere, Finland, August 97, Vol.2, pp 255-260.

[Silva,97a]- Silva, Jonny C. and Cheung, H.F.. Computer Support Environment Applying Object-Oriented Methods for Hydraulic Systems Design, Lancaster International Workshop on Engineering Design Computer Aided Conceptual Design, CACD'97, 14 - 15 April,1997.

[Silva,97b]- Silva, Jonny C. and Dawson, David. The development of an expert system for hydraulic systems design focusing on Concurrent Engineering Aspects. ICED97, Vol.1, pp 271-276.

[Silva,97c]- Silva, Jonny C. and Dawson, David. Knowledge Acquisition In The Context Of Hydraulic Systems Design- A Case Study. To be presented at the 10th Bath International Fluid Power Workshop. Bath University, 10-12 September 97.

[Turchin,92]- Turchin, Valentin. http://pespmc1.vub.ac.be/AUG1992nodes/T-AGENT.HTML.

[Tanskanen,97]- Tanskanen, Kari. On-Site Component Selection Support using WWW. ICED97, Vol.2, pp 243-248.

BEST COPY AVAILABLE

A Simple Model for Adaptive Courseware Navigation

D. Pilar da Silva, R. Van Durm, K. Hendrikx, E. Duval, H. Olivié
Katholieke Universiteit Leuven - Departement Computerwetenschappen - Leuven/BELGIUM
E-mail: {Denise.PilarDaSilva, Rafael.VanDurm, Koen.Hendrikx, Erik.Duval, Olivie}@cs.kuleuven.ac.be

1 Introduction

This paper presents ongoing research on creating adaptive background material for a last year university course on 'multimedia modeling and programming', hereafter called the *course*. The course is organized around concepts, which are explained by documents. The documents, and linking information are stored in a database. Concepts have explicit relationships with documents and with other concepts. Each document has an associated level of difficulty. The student is guided towards appropriate documents based on information about his/her knowledge of each concept.

2 The Course

A primary concern with educational hypertext is the definition of an appropriate structure so that a student can easily and naturally find the most relevant information depending on his/her needs.

Irrelevant information and links overload their working memories and screen. In order to overcome this problem it is possible to rely on information about particular users (represented in a user model) and then adapt the content (*adaptive presentation*) and/or the links to be presented to that user (*adaptive navigation support*) [Brusilovsky 96]. For the moment our approach is mainly concerned with *adaptive navigation support.*

We rely on *typed links* to represent the structure of the course which is organized around concepts, that are explained by a set of *documents.* Links between concepts represent semantic interrelationships. At present we are considering only two link types: *is_prereq_of* and *is_specialized_by*, but we plan to augment the model in the future with other link types such as *is_related_to*, *is_similar_to* and *contrasts_with.*

The *documents* are multimedia objects, such as text segments, static figures or interactive demonstrations. The URL of these documents is stored in the database. Each *document* has an associated *level of difficulty* with respect to the concept it belongs to, which varies from *0* to *99*, where a higher weight means '*more complex*'.

3 Student Model and Adaptive Navigation

The user model currently deals only with knowledge about each concept. We initialize student knowledge (or *level of expertise*) for a particular user as *0* for every concept at the beginning [Calvi 97] and update this value after the student has visited a document related to the concept. We are aware that this is a rather naive approach, but we have adopted it for the time being as it provides a simple yet flexible basis for experimentation.

The level of expertise determines the documents available to a student. Basic concepts, that have no prerequisites, can be accessed by a new student. Acquiring these *basic concepts* enables the student toconsult documents related to more advanced concepts.

The current knowledge *k* of a student *s* about a particular concept *c* can be described as *k(s,c)* *with* $0 \quad k(s,c) \quad 99$.

Associated to the *is_prereq_of* relationship between two concepts, there is a *threshold* , that represents the minimal level of expertise a student must attain in order to access the more advanced concept.

The *set of relevant concepts* (RC_s) for a student can be defined by the rule:

$$RC_s = \{ c \mid basic(c) \qquad c' : (c' \; is\_prereq\_of \; c) \quad (k(s,c) \qquad c', c) \quad \}$$

This means that a student can access basic concepts, as well as concepts whose prerequisites he/she masters sufficiently well.

The documents accessible to a particular student are those that belong to the *set of relevant documents* (RD_s), defined as follows:

$$RD_s = \{\ d \mid [explains(d,c)] \quad [c \quad RC_s] \quad [(k(s,c) - \qquad diff\_level(d) \quad (k(s,c) + \quad)] \quad \}$$

Here, *d* represents a document and is a constant. The above expression means that the relevant documents are those that explain a relevant concept with an appropriate difficulty level (as defined by).

When a student visits a document, his/her level of expertise is updated in the following way:

```
if (d   $RD_s$   $k(s,c) < diff\_level(d)$)
 then    $k(s,c)$   $diff\_level(d)$
```

The work described in [Signore 97] explores similar ideas in a non-educational application, where the user has more control (for instance with respect to the threshold value).

4 Current Status and Future Work

Crrently, all information is stored in a database implemented with Access DBMS and dynamically translated into HTML. There are several approaches to do this, varying from more traditional CGI to server-side includes and newer technologies like *servlets.* At the moment, we rely on the Active Server Page technology developed by Microsoft, but we will use a new approach we developed, where a Java based web server instantiates the necessary classes from a persistence layer [Hendrikx, Duval & Olivié 97]. The application will be tested with students at K.U.Leuven at the beginning of next academic year (between October and December 1997).

In the near future, we want to elaborate the user model by including some cognitive characteristics which are relevant for learning processes, like for instance cognitive style [Höök 96, Wilkinson 97] or reasoning abilities. We will enrich the prototype with *stereotypes* that will also be part of the user model. Thus, a student will only see the links that are associated with the stereotype representing his/her profile.

Another enhancement we are currently implementing is the dynamic drawing of local overview diagrams or concept maps that show the immediate neighborhood in order to help minimize cognitive overhead.

Tests can also be included as simple documents, making it possible to evaluate any previous knowledge a student could have about the subject in order to allow him/her to skip known concepts. We can, as well, use those tests to assess and update the knowledge a student actually acquired when following the course.

REFERENCES

[Bru 96] Brusilovsky, P., ***Methods and techniques of adaptive hypermedia***. User Modeling and User-Adapted Interaction Journal (UMUAI) v. 6 pp. 87-129. 1996.

[Calvi 97] Calvi, L. and De Bra, P. ***Improving the Usability of Hypertext Courseware through Adaptive Linking***. Proc. of the 8th ACM Conference on Hypertext. pp. 224-5. April 1997.

[Davis 95] Davis, H. ***To Embed or Not to Embed…*** Communications of the ACM, 38(8), pp. 108-9. August 1995.

[Hendrikx, Duval & Olivié 97] Hendrikx, K., Duval, E. & Olivié, H. ***HOME - The Next Generation***. WebNet'97 World Conference. Toronto, Canada. Oct 31 - Nov 5 1997. Accepted.

[Höök 96] Höök, K. ***A Glass Box Approach to Adaptive Hypermedia***. Ph.D. Diss. Stockholm Univ. October 1996.

[Signore 97] Signore, O. *et al.* ***Tailoring Web Pages to Users' Needs***. Proc. of the workshop "Adaptive Systems and User Modeling on the WWW" - 6th International. Conference on User Modeling. June 1997. http://zeus.gmd.de/UM97/proceedings.html

[Wilkinson 97] Wilkinson, S. ***Document Structure, Individual Differences and the Learning Process***. Poster at the 8th ACM Conference on Hypertext. April 1997.

ACKNOWLEDGEMENTS

The financial support of CNPq/Brazil for D. Pilar da Silva's scholarship is gratefully acknowledged. The SOCRATES program of the European Comission provided financial support for the elaboration of the courseware, in the framework of the EONT project (http://hyperg.softlab.ntua.gr/EONT).

Virtual Institute Center for the advancement of two-year to four-year transfer

JR Garcia de Alba
TX A&M System AMP Program/TEES
015 WERC, TAMU Campus
College Station, TX 77843-3405
Ph. (409) 862-4375 or 4315 or 1070

The Concept

The Texas A&M Alliances for Minority Participation (AMP) is a National Science Foundation supported program with one mission: To substantially increase the quantity and quality of minority students receiving baccalaureate, master and doctorate degrees in science, mathematics, engineering and technology (SMET). At this stage of the AMP program's development, it is important that information about its achievements be effectively disseminated so that people all over the country can benefit from this national investment.

The Texas AMP Virtual Center for Transfer and Articulation (VITA) will be created to help satisfy this need with regard to community college to four-year college transfer: a topic of particular relevance to minority populations who make up almost half of the national community college students. VITA will be used for disseminating educational policy, practice and reform information on two-year to four-year college transfer and articulation[1] through the WWW especially as it pertains to the sciences, engineering and mathematics. It will provide users such as other AMPS, academia, government, industry, and organizations with free, easy, fast and friendly access to national and specialized information. The VITA web site will include search engines, on-line forums, and data collection instruments just to name a few of the services to be provided. Other examples are:

- ✓ Information summaries on current practices, bibliographies, lists and contact information on human resources, workshops, conferences, projects and web-sites related to two year to four year college transfer and articulation.
- ✓ Customized support through workshops and one-on-one consultations to other AMPS, and other institutions of higher education.
- ✓ Primary and Secondary data collection assembly, analysis and reporting.

The Technology

---

[1] Articulation is the process of evaluating courses to determine whether a particular course offered at a junior college is comparable to, or acceptable in lieu of, a corresponding course at a four-year University.

Windows NT Server 4.0 with IIS (Internet Information Server 3.0) will be the operating system that will be utilized on the dedicated server. Frontpage 97 will be used to design the web pages and achieve a custom interface, C++ to write the common gateway interface (CGI) program for interfacing external applications with information servers such as HTTP or Web Servers, and Microsoft Access to store and manage the data. The VITA Web Site will be an interactive application capable of real-time execution and dynamic output. Multiple sets of qualitative and quantitative data will be linked in a large relational database and queries to particular content areas will be routed through the entire database extracting related data and assembling them in user friendly reports.

In Brief

It is VITA's intention to contribute to the improvement of transfer of community college students into four-year institutions by providing organized and customized information about transfer and articulation through the WWW. It is our assumption that knowledge exposure on this subject will lead to improved transfer rates and transfer success which will in turn increase the number of minority students obtaining college degrees in science, mathematics, engineering and technology. A great effort is required but the ideal is worth it: to help open the door of a better future for a quarter of our nation's population with the key of a good education.

An Adaptive Multi-Agent Architecture for the ProFusion* Meta Search System

Yizhong Fan
Sprint Corporation
yizhong.fan@mail.sprint.com

Susan Gauch
Department of Electrical Engineering and Computer Science
University of Kansas
sgauch@eecs.ukans.edu

*http://www.designlab.ukans.edu/profusion/

The goal of this project is to develop an intelligent, adaptive Web search tool. Our work is based on ProFusion, a Web meta-search engine developed at the University of Kansas. ProFusion analyzes incoming queries, categorizes them, and automatically picks the best search engines for the query based on a priori knowledge (confidence factors) which represents the suitability of each search engine for each category. It uses these confidence factors to merge the search results into a re-weight list of the returned documents, removes duplicates and, optionally, broken links and presents the final rank-ordered list to the user. The main goals of the current research are to 1) provide ProFusion with a multi-agent architecture which is easier to extend, maintain and distribute and 2) to include automatic adaptation algorithms to replace the hard-coded a priori knowledge.

The multi-agent system consists of four different types of agents, namely, a dispatch agent, a search agent, a learning agent, and a guarding agent. The *dispatch agent* communicates with the user and then dispatches queries to the search agent and the learning agent. The *search agent* interacts with the underlying search engines and is responsible for reporting search results, confidence factors, and time-out values of the underlying search engines to the dispatch agent, as well as invoking the guarding agent when necessary. The *learning agent* is in charge of the learning and development of the underlying search engines, in particular adjusting confidence factors. The *guarding agent* is invoked when a search engine is down and it is responsible for preventing the dispatch of future queries to a non-responsive search engine as well as detecting when the search engine is back online. Figure 1 shows the control flow and intercommunication between agents in the ProFusion system.

Our multi-agent architecture demonstrates various desirable agent characteristics [Mae 1994] including: task-oriented modules, task-specific solutions, de-emphasized representations, decentralized control structure, and learning and development. The search agent, learning agent, and guarding agent each consists of a set of 6 identical competence modules, each of which is responsible for one of 6 underlying search engines (task-oriented modules). These competence modules are self-contained black boxes which handle all the representation, computation, "reasoning", and execution that is necessary for its particular search engine. Although all 6 competence modules for each of the 3 agents are implemented using identical code, each uses its own local configuration and knowledge files to achieve its competence (task-oriented competence). In other words, there is no central representation shared by the several modules. Instead, every task-oriented module represents locally whatever it needs to operate autonomously. The localized representations of different modules are not related (de-emphasized representations).

Figure 2 illustrates the ProFusion multi-agent system architecture view. This architecture is highly distributed and decentralized. Each search engine keeps its competence modules and local representations in a separate directory. Except for the dispatch agent, all of the competence modules of the search agent, learning agent, and guarding agent operate in parallel. None of the modules is "in control" of other modules (decentralized control structure). Because of this distributed operation, the new system is able to react quickly to changes in the environment and make the corresponding adjustments.

The adjustments are made by the learning agent which uses adaptation algorithms. The new ProFusion adapts to changes in search engine's performance, to changes in search engine's response time, and to changes in search engine's result formats. The adaptation to performance is achieved by observing user behavior to provide feedback which dynamically changes the performance knowledge base, the adaptation to response time is achieved by using dynamically changing time-out values, and the adaptation to result formats is achieved by using a dynamic extraction pattern, or in other words, a parser.

With this adaptive multi-agent architecture, the ProFusion system is now more competitive in the dynamic Web environment since it automatically adjusts to changes in its environment. ProFusion is also much easier to maintain and extend because it no longer requires *a priori* knowledge of a new search engine's confidence factors for each category (this will be determined by the learning agent). Since the search agent incorporates a parser, no more custom code is needed for extracting the search results, only a description of the language the search engine currently speaks.

References

[Mae 1994]. Pattie Maes. (1994). Modeling Adaptive Autonomous Agents. *Artificial Life Journal*, edited by C. Langton, Vol. 1, No. 1 & 2, pp. 135-162, MIT Press.

Using Active Filters to Improve Foreign Language Instruction

By Virginia M. Fichera, Doug Lea, and Joseph Grieco,
Languages Across the Curriculum Project,
State University of New York at Oswego, Oswego, NY, 13126

Background

Use of World Wide Web resources in higher education has been primarily focussed on providing either passive research resources or Java applets for specific content interactions. However, the need often arises to provide interpretation, assistance, and commentary for existing web resources, so that students can focus on the educational objectives surrounding an assignment to view a given web page. While this need surely arises in many educational contexts, we first encountered it when using the web as a vehicle for teaching foreign languages. SUNY-Oswego's Languages Across the Curriculum (LAC) Project is an interdisciplinary effort funded by the SUNY Office of Educational Technology to internationalize the curriculum. The project facilitates student access to Web resources from many countries and in many languages in support of selected courses from across the curriculum. Faculty and students from various disciplines (e.g., economics, environmental science, history, marketing, Native American Studies) team up with faculty and students in modern languages, computer and information sciences, and graphic design to create multilingual course Web sites which organize and guide students through foreign Web sites for each of the selected courses.

The use of existing content-oriented foreign-language web sites as learning resources is an attractive way to help teach foreign languages and cultures while simultaneously providing information to students surrounding the domain at hand. For example, it is far superior to direct students to German web sites discussing environmental policies than for American instructors to prepare their own materials. There are, however, disadvantages: Because these sites are very real, often constantly updated, and not under instructor control, they make no accommodation to students who are reading them in part to become more familiar with a given language or culture. We identified three problems:

- The use of rare foreign terms, difficult grammatical constructions and the like that can make pages impossible to understand without a bit of translation or annotation.
- The lack of guidance about exactly why a certain URL was assigned to be visited.
- The lack of embedded links that help students traverse through web pages that are related for the purposes of their assignments

One way to deal with these problems is to locally copy foreign pages and customize them by hand. However, this can run against copyright conventions, and presents a never-ending obligation to update local versions when the originals change.

Active Filters

The logistical and technical problems surrounding the need to provide guidance to students can be solved by devising interpositioning tools based on *active filters*. Active filters use a customized HTTP server that intercepts URL requests and returns not only the requested page, but also any kind of assistance that is available for that page. There are several ways to implement such a tool. Our current version is based on Meta-HTML, a web server that interprets a lisp-like programming language embedded within specially written web pages. Another in-progress version uses Java Servlets to the same effect. (These implementations are freely available from the authors.) Across implementations, the basic strategy is as follows:

- Users must start out from a specially crafted web page. This page encodes links as special server directives. However, all pages presented from that point on during a session are automatically converted to use encoded links. Because all processing is performed by the proxy HTTP server, the resulting pages can be viewed in any browser.
- Each link within a viewed HTML page is encoded as a directive to the local server to fetch that page itself, to encode its links, and to analyze it for content before sending it to to the client browser. Although the content could be analyzed in just about any fashion, we currently support only two techniques, which have sufficed for our purposes:
    - Site-specific, just based on the URL itself.
    - Term-specific, triggered by the presence of predefined keywords found anywhere in the

document..

- All available assistance for the page is provided via links to *local* web pages; often just small ones providing a few annotations, a concise term definition, or a suggestion about where to go next. These links are issued separately from the main content of the page, in either of two ways:
    - As a list of labeled links appended to the bottom of the original page.
    - As a Java-based popup menu.

  We found it necessary to use such unobtrusive methods. Inserting links or menus into the foreign pages themselves nearly always disrupts the intended formatting of the original document.
- The control information (URLs, keywords and associated links) are maintained in ordinary local files that can be edited by instructors whenever they write new annotations. To simplify this process further, we are adding an HTML form-based utility for use by instructors.

Conclusions

While they are intrinsically special to a given domain and purpose, we have found active filters to be relatively easy to program and maintain using either Meta-HTML or Java. The only real complaint we have had is that since browsers are unable to cache manufactured pages, the delays encountered when fetching fresh copies from overseas on each access are sometimes too long. This problem could be addressed by having the filter itself cache pages. Active filter tools offer new prospects for interactions both between the student and the Web resource and between the student and the instructor, and help in the transformation of the role of the teacher ``from sage on the stage to guide on the side". Instructors virtually accompany students to the foreign site, providing students with assistance tailored to the filtered site. While we have found active filters to be a necessity for assisting students with Web research involving foreign languages, their uses are obviously extendible to any educational Web site, providing instructor-tailored materials which provide assistance surrounding another site without altering or copying it.

Last modified: Sun Sep 7 13:33:42 EDT 1997

Instructional Design Issues for Web-based Instruction: Teaching an Old Model New Tricks

Kristina A. Fritts
University of Arkansas, Fayetteville
Educational Technology, Grad Ed Bldg. 339
Fayetteville, AR., USA 72701
501-575-5100
kfritts@comp.uark.edu

Cheryl Krawchuk
University of Arkansas, Fayetteville
Educational Technology, Grad Ed Bldg. 314
Fayetteville, AR., USA 72701
501-575-5111
krawchuk@comp.uark.edu

Introduction

Throughout educational history changes in theory as well as technological advances have influenced the evolution of instructional design models. Traditional instructional design models, such as those developed by [Dick and Carey 1978] and [Merrill 1983], provide a basis for effective and efficient instructional design in a variety of settings. However, the incorporation of computers into the educational arena has produced a new set of design issues with regard to the creation of effective instruction. These issues have led to the development of instructional design models geared specifically toward computer-based instruction. The Rapid Prototyping Model [Tripp & Bichelmeyer 1990] and [Hannafin and Peck 1988]'s CAI design model are illustrations of this trend.

In keeping with the advances of technology, the recent explosion of web-based instruction is currently demanding that designers take yet another look at the instructional design process. As technology transforms traditional educational settings into global electronic classrooms, instructional design models must adapt their current components to focus on the unique issues encountered in the Internet environment. This paper explores some of the unique design issues of web-based instruction within the context of four basic phases of traditional instructional design models; 1) Analysis, 2) Design and Development, 3) Implementation, and 4) Evaluation. The following presents a brief overview of the aforementioned issues.

Analysis

The initial task in the design of any form of instruction, is to perform an analysis of the 1) learner, 2) environment, and 3) content. Creating web-based instruction alters the focus of this analysis somewhat from that of traditional instruction. For example, the analysis of the learner takes on several new dimensions due to the potential number of learners, their competencies, and learning styles. Rather than obtaining precise information about a very specific and identified group of learners, the designer must now take into consideration the possibility that millions of unidentified learners will be involved. Likewise, distinctions occur between traditional instruction and web-based instruction when analyzing the environment. In web-based instruction, attention is shifted from the physical setting to specific hardware and software needs of the instruction and learner population. Lastly, content must be analyzed in the context of an Internet format, as opposed to a traditional setting.

Design & Development

The second general phase of the instructional design process involves the design and development of the instructional materials. During the design stage, specific instructional methods and activities are identified as well as the objectives for the instruction. While traditional settings allow for a variety of instructional methods and activities, designing instruction via the web immediately imposes limitations on the types of methods and activities applicable. Designers must be aware of these limitations and employ alternative methods and activities that are equally effective.

The development phase involves the actual generation of the instructional materials. For traditional instruction, this may include the creation of handouts, lecture outlines, instructional props, etc. However, for web-based instruction, this phase demands much more of the developer due to the extreme technical knowledge that is required to produce web-based instruction. Additionally, materials may not be as readily available as they are for traditional instruction due to the "newness" of this form of instruction, and strictly enforced copyright laws surrounding graphics and software.

Implementation

Once materials have been designed and developed, the instruction must be implemented. In traditional instruction, this generally involves gathering the learners at a specified place and at a designated time. Web-based instruction, however, is not limited by time or space. Rather, implementation is limited by computer access and the maintenance of the instructional materials. For example, learners must have access privileges to the learning environment as well as the appropriate hardware/software in order to view the instruction. Likewise, the instructional site must be maintained by updating the links and debugging programming errors to ensure constant availability for the learner.

Evaluation

The final phase of the instructional design process involves the evaluation of the instruction as well as the learner. Various evaluation methods are employed to measure the effectiveness of traditional instruction, and many of these methods may also be used to evaluate web-based instruction. However, evaluation of the learner in an Internet environment is dependent upon the purpose of the instruction itself. Designers must differentiate between open forums dedicated to self-improvement, and web-based instruction affiliated with an educational institution in which tuition and academic credit are involved. On-line testing differs greatly from traditional evaluation with regard to the issues of accountability and security. Designers must ensure that the evaluation instrument is necessary and appropriate for the content presented.

Conclusion

In conclusion, this paper has supplied a brief look at areas in which traditional design methods may need to be adapted to accommodate the design of web-based instruction. It is our belief that we can contribute to current literature in this new area of instructional design by using traditional design models as the basis for exploring new issues facing web-based instruction.

[Dick and Carey 1978] Dick, W., & Carey, L. (1978). The systematic design of instruction. New York: Harper Collins Publishing.

[Hannafin and Peck 1988] Hannafin, M. J., & Peck, K. L. (1988). *The design, development, and evaluation of instructional software*. New York: Macmillan Publishing Company.

[Merrill 1983] Merrill, D. (1983). Component display theory. In C. M. Reigeluth (Ed.), *Instructional design theories and models*. Hillsdale, NJ: Lawrence Erlbaum Associates.

[Tripp & Bichelmeyer 1990] Tripp, S., & Bichelmeyer, B. (1990). Rapid prototyping: An alternative instructional design strategy. *Educational Technology Research and Development*, 38(1), 31-44.

A Web-Based Virtual Reality Navigation Research Tool

Greg Furness
Caterpillar, Inc., USA
e-mail: gfurness@mtco.com

Joaquin Vila, Ph.D.
Applied Computer Science, Illinois State University, USA
e-mail: javila@rs6000.cmp.ilstu.edu

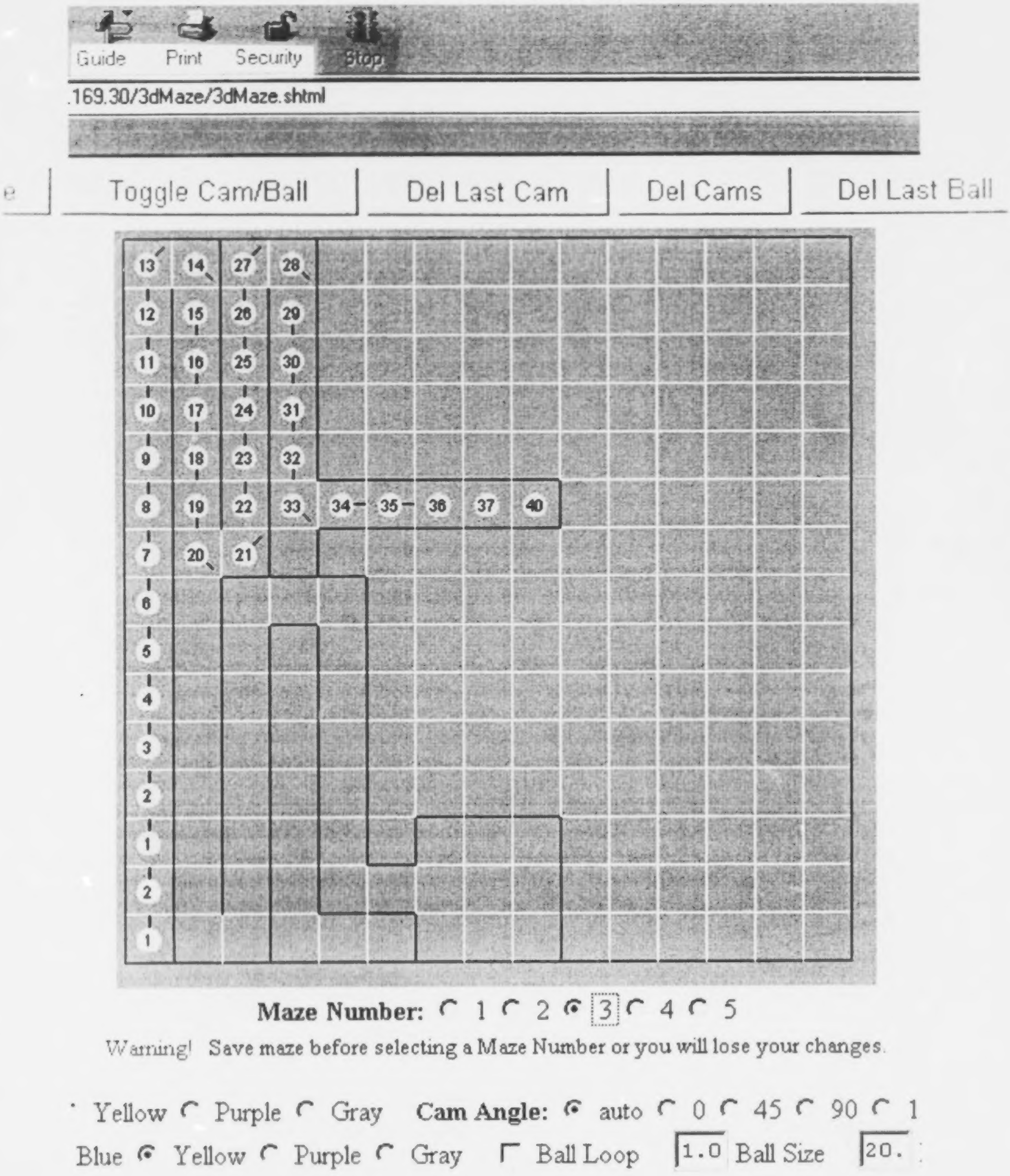
Barbara Beccue, Ph.D.
Applied Computer Science, Illinois State University, USA
e-mail: bbeccue@rs6000.cmp.ilstu.edu

This paper describes a web-based navigation tool used to conduct research into factors that affect navigational decisions in a virtual reality environment. The tool allows the design and exploration of three-dimensional worlds and provides for navigational tracking. The structure of the virtual worlds is confined to 3D mazes. The tool enables researchers to customize experiments that allow for the testing of subject's navigational and spatial behavior.

The **VR Navigation Research Tool** application consists primarily of three web pages - the **VRML 2.0 Maze Builder** page, the **Survey** page and the **Experiment** page. These web pages were constructed using state of the art web technologies including Java, JavaScript, LiveConnect classes for Java/JavaScript interconnectivity, VRML 2.0, VRML EAI (External Authoring Interface) classes for Java/VRML 2.0 interconnectivity, HTML 3.2, CGI/Perl, Netscape cookies, and OmniHTTPd SSI (Server Side Include).

The original basis of **VRML 2.0 Maze Builder** was derived from Brian Nenninger's **3D VRML Maze Builder** (*http://www.vt.edu:10021/B/bwn/3dMaze/*). The **VRML 2.0 Maze Builder** allows the researcher to easily construct a polygon-based 3D maze and provide that maze with properties and simple animations that are conducive to studying factors that affect navigation in a 3D environment. The VRML 2.0 maze that is built by this application is generated from data that the research designer inputs by pointing and clicking on a 2D grid-representation of the 3D space.

The 15x15 grid represents an overhead view of the space. The grid's x-axis corresponds to the VRML maze's x-axis, and the grid's y-axis corresponds to the VRML maze's z-axis. The lines that make up the grid represent potential walls in the maze. The walls are delineated by clicking on the lines. (See Figure 1.)



The **Survey** page collects demographic information from each subject who participates in the study. Upon completing the survey, the subject proceeds by loading the **Experiment** page. Here, the subject follows the written directions for the particular experiment and navigates the 3D maze to complete the assigned task. The movements of the user during the experiment are tracked, recorded and saved. Both the **Survey** and the **Experiment** pages can be accessed from the **VRML 2.0 Maze Builder**.

The coordinates contained in the results file are based on the 15x15 maze grid. The grid's origin of 1,1 starts in the lower left-hand corner. The locations viewed during the initial walkthrough are the first coordinates recorded; this helps identify the particular experiment (the maze definition is also included in the results file for identification purposes). Subsequent movements by the subject are then added to those initial coordinates. The current Universal Time Code (UTC) is recorded with each coordinate to indicate when the movements were made. The Start button must be clicked to start the navigation tracking. The Finish button must be clicked to save the tracking results.

The **3D Navigation Study** application was designed and tested on a Pentium 200 running Windows 95. The web pages should be accessible from *any* platform running Netscape Navigator 3.01 and a VRML 2.0 browser with EAI capabilities.

Dynamic Construction of Geographical Hyperdocuments

Mauro Gaio, Maurice Szmurlo, Jacques Madelaine.
GREYC - CNRS UPRESA 6072
University of Caen, Esplanade de la Paix, 14032 Caen Cedex, France
{Mauro.Gaio, Maurice.Szmurlo, Jacques.Madelaine}@info.unicaen.fr

Introduction

The recent development of the Internet allows anyone to easily publish documents on the World Wide Web. Unfortunately, most of the time these documents are simply electronic clones of their "paper" counterparts both in their structure and in the way they are linked to each other. Moreover, construction of "multi-article" documents relies on 1) centralization of all the the papers once they have been written, and 2) on their transcription into the electronic format. In other words, the new possibilities offered by network communication and (almost) instantaneous access to information are almost never used. This problem is even enhanced when the elaboration of the document requires skill and knowledge from geographically distant authors. These new possibilities can improve both the editorial cycle and the reading process, that is the way the authors submit their articles, the way these papers are reviewed, and the way the reader will access the final document.

This position paper presents the current state of our analysis of a tool which allows dynamic and collaborative construction of a document organized in themes, sub-themes, etc. each of these composed of highly specialized and independant papers. As a practical example of such a document, we work with geographical documents, and more specifically geographical atlases. The result of the construction, process we call for "editorial cycle" and which will be discussed in the first section, is a hyper document that can be consulted through the WEB. We also will show in the second section how the electronic medium allows several reading strategies based both on the wishes and the knowledge level of the potential reader.

Editorial Cycle

From the authors point of view, the paper is added onto the documentserver remotely by being inserted directly into the logical structure of the document which had been defined by the editorial team. Once this is done, the article becomes immediately available to the authors community for review and comments. If the document is multi-language, as it is the case in our experimental document, the different translations of the same paper may also be uploaded.

Beside the possibilitily for the authors to work remotely, network usage offers several advantages. Since the authors will not insert their papers simultaneously, the document is most of the time incomplete. However, this gives a partial view of the whole document as it is being constructed and allows early redundancy tracking. Moreover, reviewing can also be performed electronically using an annotation system based on short notes put aside of the part of the text they comment [Schickler *et al.* 96] ("Post-It"-like notes). These "in-the-context-comments" are made available with the paper so that other authors are aware of the remarks that have already be done. Moreover, as the comments are in the context of the part they comment, reviewing is much more efficient that smail, email, or fax exchange between authors and reviewers.

Secondly, the electronic medium offers a more powerful inter-article linking framework. Rather than linking a paper to a set of predefined target papers (that may not have already been inserted), the author will specify parts of his text or graphic illustrations as outgoing links defined by output descriptors (OD). Following such a link results in a search among the existing papers and retrieval of articles that match these descriptors. This assumes that during insertion, the papers input descriptors (ID) have been defined in order to allow their indexation. IDs and ODs are for the moment limited to keywords (sorted in a predefined set). The choice of IDs can be done manually by the author or by the editorial team, or automatically by the system (using full text indexing engine), or any combination of these possibilities.

The third advatage of the electronic medium over the classical paper medium is the possibility to create several "tables of contents", that is several logical structures for a given set of papers, thus, providing the readers with several "points of view". Authors have to insert their papers only in the predefined logical structure. The others logical structures are defined by the editorial team once all the papers have been reviewed. This enforces the readibility of the whole document.

Reading Strategies

The advantages for the readers are first of all several reading strategies. Firstly, the traditional, sequential "paper like" browsing (next page, previous page, next chapter, etc. based on the logical structure of the document) is of course available. As the document offers several logical structures, this first strategy is therefore enhanced.

As the papers are linked together in a dynamic way, the target of a link may be a set of several papers, which is unusual in traditional hyperdocuments. The problem for the user is to choose the next paper to read. The intermediate page that proposes all the possible targets has therefore to also include some more informations about each target, as for example an abstract. However, for a non-specialist readers, this information may be not sufficient. Therefore, the editorial cycle should include a phase where experts of the domain will propose the prefered targets for each of the source articles. This defines an "expert reading strategy".

Thirdly, since we are using an electronic medium, the graphical illustrations (statistical charts, maps, etc.) may gain a *à la carte* feature. Indeed, most of such illustrations are very synthetic, containing many symbols, colors representing different layers, etc. and are therefore difficult to read, even in their paper format. Therefore, we propose to offer to the user the possibility to decompose an illustration into its basic layers, and allows the viewing by selecting the layers of interest. This feature can be obtained if, rather than storing the image of an illutration, we store the methods and the data that were used for its construction [Szmurlo *et al.* 96]. Obtaining the methods and the data would put many constraints on the authors work as they would need to use the same graphical, analysis, cartographic, etc. tools.

Finally, the reader can, of course, perform a search based on keywords or pieces of sentences. This "free reading" is offered "for free" since all papers are described by their ODs and as a search engine has been implemented for link computation.

Technical Issues and Conclusion

We are currently working on the implementation of such a collaborative system with a subset of the features defined above. Its architecture is client/server: a central document server will contain all the documents, while the clients (authors and readers) will use their favorite WEB browser in order to access the papers. The author part uses a Java applet in order to perform uploading of the papers. The papers are in RTF format which allows the authors to use their favorite text processing application as well as to integrate text and images into a single file. Once uploaded on the server, a paper will be transformed into HTML. Illustrations are extracted and stored in GIF format. Texts, illustrations, ODs, IDs, and other information about the author are eventually stored in a relational database which allows dynamicity and easies document management. Reading is performed by calling a CGI script which accesses the database.

The system described in this paper is also a very interesting experimental tool for information acquisition and retrivial. The first point is the design of a robust seach engine which will link together papers. The current version is limited to keywords and is purely syntaxic. It will envolve in a syntaxico-semantic analyser wich will be able to use ODs and IDs defined by pieces of sentences. As we work on real material (our document is a real geographical atlas), and since it is possible to track the readers wandering in the document, an interesting research direction would be to define user profiles and propose to each user personalized reading advises.

References

[Schickler *et al.* 96] Schickler, M.A., Mazer, M.S., and Brooks, C. (1996). Pan-browser support for annotation and other meta-information on the World Wide Web. *5th Int. WWW Conference, 1996*, 1063-1074.

[Szmurlo *et al.* 96] Szmurlo, M., Crémilleux, B., Gaio, M., and Madelaine, J. (1996). Toward an efficient cross-fertilization of multiple information sources, *Workshop on Knowledge Representation (W20), ECAI'96*, 1996, Budapest, Hungary.

Targeting User Needs with Customized Web Views

Kathryn F. Gates
The Mississippi Center for Supercomputing Research
University of Mississippi
USA
cckathy@olemiss.edu

The recent growth of the World Wide Web (WWW) has been staggering. By providing almost universal accessibility and platform independence, with appealing and easy-to-use client interfaces, the WWW has become a highly successful choice for delivering information. Yet, WWW technology is new and has a number of limitations. For example, "surfing the Web" can be a time-consuming and fruitless process. Where will a given link go? Does the user have any control over the navigation of Web pages or over the content that is displayed?

Over a decade of research on hypermedia applications offers promising possibilities for enhancing the WWW – the most well known hypermedia system to date. Ideally, a hypermedia application is structured to avoid information overload; the characteristics, motivations, and needs of different individuals can be taken into account. Bieber and Vitali [Bieber and Vitali 1997] offer the following advice: "Web Environments should not overwhelm users by providing too many options. Web environments should include filtering mechanisms to present only the most relevant links, based on the users' current goals."

The Application Domain: An On-line Reference Guide

This work attempts to achieve customized Web views in an Intranet setting. The application domain will be a Web-accessible reference guide for the Mississippi Center for Supercomputing Research (MCSR) user community. Each Web page in the on-line guide will be dynamically constructed and delivered to the end user based on his or her interests, experience and current goals, with the objective being to present those components which are most relevant rather than requiring the user to sort through the entire collection. Because the MCSR user community is diverse, including novice computer and network users as well as experienced programmers, the on-line reference guide is well suited as the application domain.

An Object Server to Deliver Reference Guide Content

Content for the reference guide will come from text files, HTTP documents and relational databases distributed across an Intranet. The Xerox Parc Inter-Language Unification System (ILU) [Janssen and Spreitzer 1997] will be used to model the content as objects and to implement a three-tier server interface to those objects. Reference guide objects will include tasks, examples, terms, consultants, knowledge base items and so on. Methods will be made available to client applications allowing them to operate on reference guide objects. For example, client applications will be able to find out if objects exist, retrieve objects, and display objects as text or HTML. ILU is a good choice for this application because it provides a handy means for building distributed client-server applications, it is nearly CORBA compliant, and its modules can be implemented in several languages, including Java. The Java support is especially desirable because JDBC will be used to access the relational databases where possible. On the other hand, parts of this application are expected to be computationally intensive may be more appropriately implemented in a highly optimized language such as C.

Indexing Reference Guide Objects with Text Analysis Tools

Text analysis tools such as the Vector Space Model [Salton and McGill 1983] provide a remarkably effective scheme for representing the content of a collection of documents which can then be used to select items related to a user's interests. In this application, text analysis tools will be applied to reference guide objects, resulting in a set of term vectors representing the reference guide content. A robot will periodically and systematically analyze the collection of reference guide objects to build the full-text analysis indices. Performance is a key issue in any interactive hypermedia application. The Vector Space Model has been chosen for this application because (a) it offers outstanding support of relevance feedback, (b) it supports ranked output and (c) implementations of the Vector Space Model can be achieved with reasonable computation time.

Accessing Reference Guide Objects from the Web Server

The reference guide Web server will be equipped with Java servlet support, and a Java servlet will act as an ILU client to retrieve and display reference guide objects. Servlets provide several advantages over CGI programs. Most significant to this application is that server connections can be established once rather than for each hit to a Web page. A servlet will parse HTML templates which contain special "<object type=refguide>" structures. When the servlet encounters one of these structures, it will parse the parameters to determine what type of object to retrieve, how many objects to retrieve, and what criteria to use to use in matching the objects. It will then consult the robot-generated indices to select the reference guide objects that best fit the current user's goals.

Modeling the User

The problem remains of constructing a query vector which aptly represents the user's current goals – or, modeling the user. Query vectors will be constructed by (1) employing user profiles, (2) augmenting user profile information with a representation of the user's current goals, and (3) providing a built-in mechanism to allow users to iteratively refine a search. User profiles will be established on the first visit and stored for future visits. The profile will be constructed by presenting the user with a brief set of questions to determine her level of expertise, interests, background, etc. Examples of questions might be: "How long have you worked with UNIX systems?", "Do you come from a scientific/engineering background?", or "What percent of your time do you spend doing Web development?". On each visit to the hypermedia application, questions will be presented to determine the user's current goals. The user might be asked to complete the sentence: "Today I am interested in finding out about …". In this manner, a distinction will be made between those interests which are temporal and those which are abiding [Oard 1995]. The user will be given opportunities to periodically update her profile to take into account her changing interests over the long term.

Oard and Marchionini [Oard and Marchionini 1996] emphasize that "using techniques that exploit the strengths of both humans and machines" enhances user satisfaction in information filtering systems. At appropriate points in the application, the user will be given an opportunity to refine the retrieval results by marking those items which most closely match his or her interests as well as those which are not relevant. This feedback will be used to adjust the weightings in subsequent query vectors.

Conclusion

Opportunities abound for using the WWW to deliver information. A pressing challenge, duly noted by the hypermedia community, is to deliver customized views so as to not overwhelm the user with irrelevant information. This work's objective is the delivery of customized views on a heterogeneous body of information to an end-user through the WWW; information filtering techniques, user modeling and object oriented methods will be used to achieve the objective.

[Bieber and Vitali 1997] Bieber, M. and Vitali, F. (1997). Toward Support for Hypermedia on the World Wide Web. *Computer*, 30 (1) , 62-70.

[Janssen and Spreitzer 1997] Janssen, B. and Spreitzer, M. (1997). *ILU Reference Manual.* ftp://ftp.parc.xerox.com/pub/ilu/2.0a10/manual-html/manual_1.html (Accessed September 4, 1997).

[Oard 1995] Oard, D. (1995). User Modeling for Information Filtering. http://www.ee.umd.edu/medlab/filter/papers/umir.html (Accessed February 22, 1997).

[Oard and Marchionini 1996] Oard, D. and Marchionini, G. A Conceptual Framework for Text Filtering. http://www.cs.umd.edu/TRs/authors/Gary_Marchionini.html (Accessed February 24, 1997).

[Salton and McGill 1983] Salton, G. and McGill, M.(1983) *Introduction to Modern Information Retrieval*, McGraw-Hill.

ROADBLOCKS AND BREAKTHROUGHS IN INTEGRATING THE WWW INTO THE UNIVERSITY CURRICULUM

Dr. Fred Herschede
Division of Business and Economics
Indiana University South Bend
South Bend, IN USA 46634
fherschede@iusb.edu

Dr. E. Marcia Sheridan
Division of Education
Indiana University South Bend
South Bend, IN USA 46634
msheridan@iusb.edu

WHY INTEGRATE THE WWW INTO THE UNIVERSITY CURRICULUM

In both undergraduate and graduate teacher education, as well as in undergraduate and graduate business and economics, it is not enough to require students to understand how to use technology. We believe that university faculty must model how to integrate technology into instruction. While our goals for our students might be somewhat different, we took this challenge seriously in designing the curricula for our courses. We believe that students need to actually experience how technology is an integral part of being a teacher or an economist and that with it we could help students learn better.

As a result we wanted to use the WWW as a way to make course material available, deliver content, find research source material, and serve as a means of communication between students and professors as well as among students engaged in collaborative learning and work.

ROADBLOCKS AND OBSTACLES IN INTEGRATING THE WWW

In our initial work two years ago, we found that while many students did not have personal computers and modems, the university had committed to providing them with access to convenient, available technology in computer labs on campus. While graduate students in business were more likely to be employed in a situation which provided personal computers with WWW, practicing teachers and undergraduate economics students were virtually without such access. As IUSB is a commuter campus, campus computer labs were often 25-50 miles away from their places of employment.

Over the two years, we have found changing degrees of expertise in using the WWW. In Education, an undergraduate requirement in Computers in Education has moved us from a situation where few if any undergraduate students know how to send email or access the WWW, to currently 100% have these skills. A similar requirement in the school of business has had the same impact on economics students.

Graduate teacher education students in a graduate class still have yet to reach the 50% mark as far as familiarity with email and the internet. How to provide this instruction as part of the course content and using course time remains a problem. As more and more are becoming familiar with these skills, some of those already knowledgeable are complaining about course instruction in these skills.

As we began to require or at least allow students to incorporate the WWW into course requirements, we wanted to limit the time required for students to productively find relevant materials on the WWW versus just surfing around. Regardless of whether students were undergraduate or graduate students, the majority of our student body work, have families, or other obligations which require an efficient use of their time.

When we began to incorporate cooperative learning requirements we wrestled with the problem of how to find time for our students to get together to work not only in class, but outside of class as well. As a commuter campus with students coming from all directions, just finding a common time to work together outside of class was a problem as well.

BREAKTHROUGHS AND TRENDS IN DEVELOPMENT IN DISCIPLINES IN INTEGRATION THE WWW

Over the past two years, all groups except graduate Teacher Education students have acquired the knowledge and expertise to send email and access the WWW either at campus computer labs, personal home computers with modems or computers available in their workplace. Practicing teachers, for the most part, do not have convenient access to email or the WWW unless they have a personal home computer.

The problem of teaching the virtual WWW novices in graduate teacher education along with the experts has been addressed in several ways. First, there are free university mini-classes for introductory computers, email, WWW, and creating web pages. Novices are being required to take beginner classes on email and the WWW if they need them. This has not be received with total happiness due to other constraints on their time. Technophobia literally silently "screams" when students hear the assignment to take these free mini-classes.

Class time is being used to introduce our syllabi, demonstrate how to do appropriate literature searches via computer, how to access the library for electronic interlibrary loan, and electronic online reserve material. This instruction is a justifiable use of class time as most students are unaware of the electronic interlibrary loan form, and since electronic online reserves has just started at IUSB, none has had prior experience with it.

When we designed course materials available on the WWW as well as creating hotlinks to research materials students needed, we found that students were more likely to put in the time necessary to overcome initial obstacles. As students were required to work collaboratively despite their geographic distance from one another, the use of the WWW became a way to overcome the distance problem particularly through the use of email. Swapping email addresses and planning ahead to combine work through utilizing the same software programs has become more common. Varying levels of expertise were minimized as experts in groups quickly brought novices up to speed in utilizing the WWW and sharing resources on valuable WWW sites.

FUTURE PLANS FOR USING THE WWW IN OUR COURSES

With the proliferation of valuable materials on the WWW, one of us is experimenting with an undergraduate education course with all materials online. Some of them are from government and government-funded agencies with valuable resources. Others are journal and chapter readings available electronically via the IUSB Online Electronic Reserves. So far the students have responded favorably as printing course reading from either their home computers via modems, or in campus computer labs appeals to their money saving tendencies.

On problem that remains with the graduate education students is to minimize their anxiety regarding requiring email and WWW mini-classes. This group has a great deal of anxiety when being told that must learn these skills in addition to other course requirements. While one student literally had an angry outburst in class and dropped the course at break time, the remaining students who needed the instruction went to their children, their school technology coordinators, and campus mini-classes for the help. Generally, their responses are grateful for being forced, and express awe at what they now can do.

Issues in Distance Education

Ali Jafari, Ph.D.

Co-Director, Associate Professor

Advanced Information Technology Laboratories

Indiana University, USA

Jafari@iupui.edu

Schools and universities are totally missing the distance education (DE) boat by ignoring the Internet and Web. Schools are spending a large amount of capital on DE technology and infrastructure with little funding being committed for content development and faculty time. Industries influence the decisions that schools make when deciding which type of DE system and technology infrastructure will best suit their needs. Schools are relying too heavily on television as the medium for delivering DE programs rather than the Internet and Web. Today, many schools are losing money by offering DE courses. Schools practice DE by connecting classrooms to classrooms, rather than connecting the homebound and working students to information. If schools don't move fast in switching their DE models to the Web environment, soon most of the DE courses will be offered on web sites with names ending with a ".COM" rather than ".EDU." These and more are the major issues with DE programs in most schools and universities in the United States and around the world, today (see note 2).

Does DE deliver what it is expected to deliver?

Virtual access, low cost, and quality should be the expectations of any DE programs. These three requirements have rarely been met in many current DE initiatives and programs. Many schools spend millions of dollars on interactive video technology to transmit a live television image of one classroom to another classroom in the same city or nearby towns. This widely implemented method offers three major limitations. First, it does not solve the virtual access issue; students still need to leave their homes or work to go to remote classrooms. They are still facing problems such as parking, traffic, time, etc. Second, the cost of this method of delivery is enormously high especially when the depreciation of the equipment and infrastructure is added into the cost. And third, the quality and effectiveness of this television-based instruction model is limited due to the passive nature of the television medium as compared to the interactive features of the Web.

Is the Internet and Web ready for use in DE?

Without doubt, the Internet and Web technology in 1997 is ready to support most of the DE initiatives. The combination of the Internet as the delivery medium and the Web as the common user interface offers the most practical, cost effective, virtual, and interactive environment to support DE. In spite of the short-term limitations of the Internet, the Internet/Web offer some unique features that satellites, ISDN, and cable television transmission cannot offer for DE. The major characteristics of Web-based DE, as compared with television-based DE are as follows: virtual access, low cost, common and easy user interface, interaction, multimedia, a digital information system, high reliability, gate keeping, live or digital library access, and expert programming support.

DE programs can be divided into two major categories; the live or synchronized mode and the on-line or asynchronized mode. Each method provides certain advantages and limitations and supports certain types of courseware. The Web supports both types of DE modes. Chat, streaming audio and streaming video is an example of the Web component supporting the synchronized DE mode. Examples of a Web-based synchronized DE course can be found at http://WebLab.iupui.edu/projects/courses.html, the CPT499 multimedia course and Cpt 299, Internet literacy course. The Web also offers the best interactive environment for on-line or asynchronized DE course delivery. An example of this mode can be found at http://weblab.iupui.edu/c101demo. This five credit hour

Chemistry 101 course offers 26 hours of streaming video with synchronized multimedia content, on-line testing, chat, and JAVA tools on a full two-way interactive multimedia Web environment between students - instructor and student - student .

How to plan for DE?

Many schools spend millions of dollars of capital on the technology infrastructure and backbone with little or no funding for DE course development and faculty time. Reengineering a course for DE Web delivery could be more time consuming and resource intensive than writing a textbook. Faculty should be given release time, be provided with resources and expertise in Web programming, media production, and instructional engineering. Web laboratories and centers for teaching and learning should be established in schools in order to provide experimental opportunities with emerging Web solutions, training, expertise and resources to faculty becoming engaged in the production of DE courses. Unfortunately, this is not happening in many schools, instead, the DE funding is allocated to some expensive DE infrastructure that forces more faculty to produce and deliver the same old "talking head" passive DE courses.

School administrators and school deans should view the development of DE programs as an institutional investment that, when complete, will serve many virtual students and generate a new line of income. This is especially true with Web-based DE courses since it can reach a completely new student market, even international, that could not be reached in the past using the traditional DE technology. Development of a DE course is similar to the development of a computer program or a textbook. The investment of time and money will be paid back over time and with the number of users served.

How to redesign a course for DE.

Redesigning a course for DE delivery is a time consuming, resource intensive, and complicated task. Generally, four different types of expertise and services are necessary to architect and produce a Web-based DE course. First, a course subject matter expert; i.e. a faculty member who has taught the course and understands the course learning objectives and requirements. Second, instructional engineering and pedagogy; i.e. an instructional designer. Third, media production; i.e. a graphics and video producer. And forth, Web programming, engineering and authoring; i.e. a Web master with programming, engineering and authoring experience. All of the above services and expertise should be directed by a "course architect" who will oversee and manage the design and possible production of the course. The course architect's role is very similar to the function of a movie director directing a movie, or an architect designing a building. Depending on the magnitude of the DE project, the course budget, and course objectives, one or several people or groups of people are needed to work in the above four categories. If one person is intended to completely design and produce a DE course, this person is then expected to have and provide all of the above four categories of expertise and services. Although some DE projects have been totally designed and produced by an individual faculty member, it is more realistic to view a DE course project as a team effort with members focusing on one or more of the above four categories.

Note 1. Ali Jafari, Ph.D. is co-director of the Indiana University Advanced Information Technology Laboratories, Chief Scientist of the Indiana University WebLab and Associate professor of Computer Technology at Indiana University - Purdue University Indianapolis.

Note 2. All of the points and concepts discussed in this paper are only the author's professional thoughts and experiences and do not necessarily reflect the campus' or the University's positions on the subject of distance education.

Note 3. More information can be found in the following web sites:

- Ali Jafari: http://www.iupui.edu/~jafari/ jafari@iupui.edu, (317) 278 2630.
- WebLab: http://WebLab.iupui.edu/
- Chemistry 101: http://WebLab.iupui.edu/c101demo/
- IUPUI: http://www.iupui.edu/

- Indiana University: http://www.indiana.edu/

A Form-based Gateway for Web Databases

Euna Jeong *and* Jane Yung-jen Hsu
Computer Science and Information Engineering
National Taiwan University
Taipei, Taiwan 106
{eajeong, yjhsu}@robot.csie.ntu.edu.tw

1. Introduction

The world-wide web has become a popular vehicle for sharing and exchanging information over the network. Web-enabled applications often support dynamic information by managing data via a database gateway. Integrating databases with the web enables information providers to create and maintain structured information, which can be accessed by the end users easily. For example, in computer-assisted distance learning, an instructor should be able to create learner-specific on-line course materials from a database of actual cases relevant to the subject. However, creating such applications requires extensive knowledge about database and web programming on the part of the instructor. This paper describes a web site database gateway tool that facilitates the process of defining, constructing, and manipulating databases via a web browser.

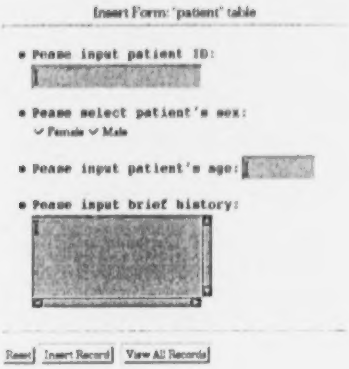
Users can create and access web databases without laborious programming. All database operations are performed by interacting with a web browser using a from-based interface. To create a new database, an information provider needs to define an appropriate schema for organizing the data. The data entry and query forms are generated automatically from simple specifications. Such a database gateway has to handle the transformations between user requests and database results [Fielding et al.96] [McCool 94]. On the one hand, when a user submits her request via a web browser, the input is encoded as form data, which needs to be decoded [Eichmann et al. 94]. The database gateway formulates the decoded user requests into SQL queries to the target databases [Bjorn and Hotaka 95]. On the other hand, query results from database management systems are translated by the gateway into properly formatted HTML documents.

Figure 1: Database schema definition

2. System Description

Consider the task of creating on-line course materials from a database of clinical cases. Organizing the cases into a database involves processes of schema design, data collection and entry. For example, a medical system administrator creates a database named *Digest* with two tables: *patient* and *case*. Figure 1 shows how she defines a specific field (e.g., *lab_report*) by identifying its name, type, length, and features. Once the data schema is available, she can design a data entry form as in Figure 2 to collect patient data. The database gateway facilitates the user to create such forms semi-automatically in two easy design steps. Figure 3(a) demonstrates the first step in which the user selects an appropriate entry method, its size and order for each field defined in the schema. Figure 3(b) shows the next step in which she specifies further details, e.g. *Help message, options, and defaults*. Such a design tool provides the advantages of constraining input data values as well as eliminating misspellings, thereby greatly simplifies the task.

Additionally, to support users to lookup and update web databases, the gateway provides template pages for the database administrator to design query forms. For example, a medical instructor may retrieve data from relevant cased in the *Digest* database through a query interface. Since the end users, e.g. medical students, usually don't have adequate training in formulating database queries properly. The query interface assists the users by providing the range and/or contents of acceptable values. A user specifies each query condition as three elements: the field name, a comparison operator, and a target. The field name is based on the current table(s). The comparison operators include *greater than, equal to, less than*, and so on.


Insert Form: "patient" table

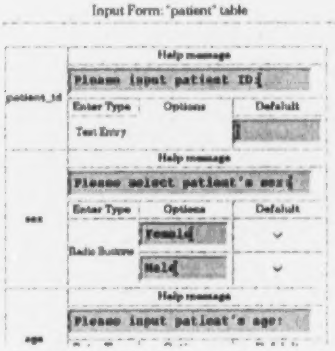
- Pease input patient ID:
- Pease select patient's sex:
  ○ Female ○ Male
- Pease input patient's age:
- Pease input brief history:

Figure 2: Input form

Figure 3(a): Design input form step1

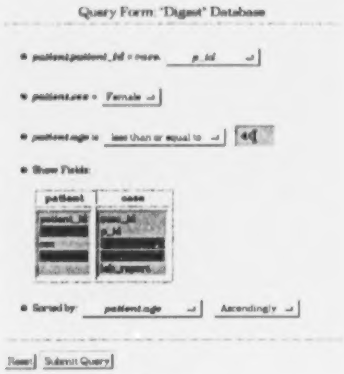
The target specifies the value string, against which each record will be compared in answering the query.

The query form in Figure 4 was created with the help of the interface as shown in Figure 5. At this point, three query types are supported. First, it can display all the values of a given field for easy selection. Second, it can display all comparison operators together with an input value area. The user can select the appropriate comparison operator and then enter a text string as the query keyword. The third query type is used for specifying relational queries between tables. In this case, a user can select the tables to be queried. By default, the gateway (in Figure 4) also provides specification of display method for the query results: fields to be displayed and the sorting order of the records.


Input Form: "patient" table

patient_id	Help message		
	Please input patient ID:		
	Enter Type	Options	Default
	Text Entry		
sex	Help message		
	Please select patient's sex:		
	Enter Type	Options	Default
	Radio Buttons	Female	○
		Male	○
age	Help message		
	Please input patient's age:		

Figure 3(b): Design input form step2

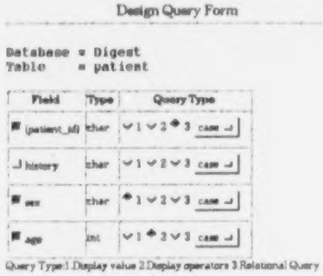

Query Form: "Digest" Database

- patient.patient_id = case. p_id
- patient.sex = Female
- patient.age is less than or equal to 40
- Show Fields:

patient	case
patient_id	case_id
	p_id
sex	
	lab_report

- Sorted by: patient.age Ascendingly

Figure 4: Query form


Design Query Form

Database = Digest
Table = patient

Field	Type	Query Type
☑ (patient_id)	char	○ 1 ○ 2 ● 3 case
☐ history	char	○ 1 ○ 2 ○ 3 case
☑ sex	char	● 1 ○ 2 ○ 3 case
☑ age	int	○ 1 ● 2 ○ 3 case

Query Type:1 Display value 2 Display operators 3 Relational Query

Figure 5: Design query form

3. Conclusion

In this research, we have implemented a database gateway tool that has the capabilities for *information access and update, database design and creation, information output format, and automatic input and query form layout design.* By using standard web browsers as a uniform interface to all database transactions, it makes the database gateway adaptable to multiple platforms and reduces the need for using a customized client application for each database. Furthermore, we offer a progressive solution to database interface design. The form design tool allows the average users to automatically generate and modify HTML form documents for data entry and query, instead of directly changing the scripts that generate the corresponding hypertext documents. Such a tool supports improved maintenance of web databases as well as error-free and efficient interface design. As a result, even those who are not familiar with programming languages, HTML tags, or SQL queries can manage databases and design user interfaces. It can also reduce the time and cost of software development for database administration over the network.

References

[Fielding et al. 96] Fielding, R., & Frystyk, H., & Berners-Lee, T., & Gettys, J., & Mogul, J. C. (1996). Hypertext Transfer Protocol, 1996, Available at http://www.w3.org/pub/WWW/Protocols/HTTP/Issues/Revs/Rev81Clean.txt.

[McCool 94] McCool, R. (1994). The Common Gateway Interface, Available at http://hoohoo.ncsa.uiuc.edu/cgi/overview.html.

[Eichmann et al. 94] Eichmann, D., & McGregor, T., & Danley, D. (1994). Integration Structured Databases Into the Web: The MORE System, *Proceedings of the First international Conference on the World-Wide Web.*

[Bjorn and Hotaka 95] Bjorn, M., & Hotaka, R. (1995). A WWW Gateway for Interactive Relational Database Management, *Proceedings of The First Australian World-Wide Web Conference.*

BEST COPY AVAILABLE

Web Server Performance and Process Model

Tae-sung Kim, Hyuck Yoo
Department of Computer Science
Korea University
E-mail {tskim,hxy}@simple.korea.ac.kr

Abstract: World Wide Web is one of the most rapidly growing fields in the Internet world. The performance of World Wide Web is critically dependent on web server so that the performance of web servers becomes more and more significant topic, yet there are not enough performance studies of web servers being done. In this paper, therefore, we investigate the performance characteristics of web servers on various hardware and software platforms. Our focus is on the process model of web servers and its relationship with the performance. We report the results of extensive benchmarks on web server performance.

Introduction

WWW[1] is rapidly growing on the Internet, and more and more people access information via WWW. The software used in WWW consists of two parts. One is web client that issues requests to web server and receives responses from the web server. The other is web server which receives requests from web clients and sends appropriate responses to the web browser. Currently, there are popular web clients such as Netscape[2] and Internet Explorer[3]. End-users have a choice from several web clients after evaluating the functionalities and performance. However, in the case of web server, because it does not directly interact with the end-user, the performance issue of web server has not been well understood. But, the performance of WWW mainly depends on web server rather than that of web client. Therefore, it is important to investigate web server performance in a systematic way, which is the topic of this paper. This paper is organized as follows. We start with explaining process models of web servers. Then the benchmark used in this paper is described. We use a de facto standard benchmark called WebStone[4] [Gene and Mark 1995]. We present the results of running the benchmark on various hardware and software configurations. The paper concludes with a summary of the analysis of the results.

Web Server Model

There are two types of process models in web servers. One is so-called multi-process model, and the other is the single-process model. In the multi-process model, each connection request from clients is handled by a separate server process. It causes the *master* server process to fork a server process in order to handle a newly arrived connection request. As shown in [Fig. 1], forking a process generally involves high operating system overhead such as context switching. An approach to avoid the overhead is to use a server pool that consists of several server processes and the processes are ready to run. However, this approach has a few potential limitations and disadvantages. The key disadvantage would be that it wastes resources by *pre-creating* servers. In contrast, the single-process model uses a single server process in order to handle multiple connection requests from clients: see [Fig. 1]. By multiplexing several file descriptors, the server can serve multiple requests simultaneously. There is no fork involved, and the operating system overhead is much reduced in connection after receiving the response. The number of WebChildren processes is configurable in WebStone. The size of a response from web server is a variable ranging from 0K to a maximum size that is configurable. The default maximum size is 200K. For the details of WebStone, please refer to [Gene and Mark 1995]. WebStone has three performance metrics. The first metric is average connection rate and computed by

---

[1] World Wide Web.
[2] Netscape is a registered trademark of Netscape Communications Corporation.
[3] Internet Explorer is a registered trademark of Microsoft Corporation.
[4] WebStone is a registered trademark of Microsoft Corporation.

$$\text{Averageconnectionrate} = \frac{\text{Total\# of connectior}}{\text{Totalexperimenttime}}$$

The second metric is average latency that consists of connection delay time and request delay time. The connection delay time is between when a connection request is issued and when the actual connection is established. The request delay time is the delay between the request transmission and the response arrival.

$$\text{Averagelatency} = \text{connectiondelaytime} + \text{requestdelaytime}$$

The third metric is average throughput that is total number of bytes divided by total experiment time. Total number of bytes include HTTP headers as well as requests and responses.

$$\text{Averagethroughput} = \frac{\text{Totaldata transmitted}}{\text{Totalexperimenttime}}$$

WebMaster waits until every WebChildren process finishes and collects the above metrics from the WebChildren.

There are some benchmark results published on the WWW [Michael 1996a] [Michael 1996a]. But, these benchmark results only present the performance in single-processor configuration and did not take into account process model of web servers.

Hardware and Software Configurations

As web server for experiments in this paper, we choose NCSA HTTPd [NCSA 1996] for multi-process model and Spyglass [SpyGlass 1996] for single-process model because they are one of the most widely used web servers. The NCSA HTTPd version is 1.4, and the Spyglass version is 1.10fc2.

	SPARCstation20	SPARCstation20
CPU	SuperSPARC	SuperSPARC * 2
Clock speed(MHz)	75	50
Memory(MB)	64	64
OS	Solaris 2.5	Solaris 2.5

Table 1: Hardware and software configuration

We use SPARCstation20 with single-processor as web client system and SPARCstation20 with dual-processor as web server system. [Tab. 1] is the detail specification of each system. By turning on and off a processor on the web server system, we test web server on single and dual processor configurations.

Benchmark Result and Analysis

The results on the single-processor configuration are in [Tab. 2]. We ran WebStone more than 20 times. It shows that Spyglass of single process model has better performance than HTTPd. Spyglass can handle 16.36 connection more than HTTPd in average and have lower latency by 5.50 second. It also transmits 0.7Mbit per second more. [Fig. 3] plots the average connection rates. The peak difference takes place at 180 clients, and Spyglass is 34 percent better than HTTPd.

Total number of clients	Average connection rate (conn/sec)		Average latency (sec)		Throughput average for all connections (Mbit/sec)	
	HTTPd 1.4	Spyglass 1.10fc2	HTTPd 1.4	Spyglass 1.10fc2	HTTPd 1.4	Spyglass 1.10fc2
30	56.85	65.20	0.5302	0.4706	3.12	3.25
60	56.83	68.88	1.1149	0.9330	2.97	3.75
90	58.30	73.08	1.7125	1.2966	3.05	3.97
120	59.08	76.50	2.2145	1.7148	3.06	4.00
150	58.68	78.05	3.2909	2.0869	3.56	4.02
180	60.05	80.57	3.5301	2.4918	3.21	4.24
210	60.07	75.12	3.9858	2.9089	3.07	4.12
240	69.67	93.02	3.7377	2.7169	3.80	4.88

Table 2: Results on single-processor configuration

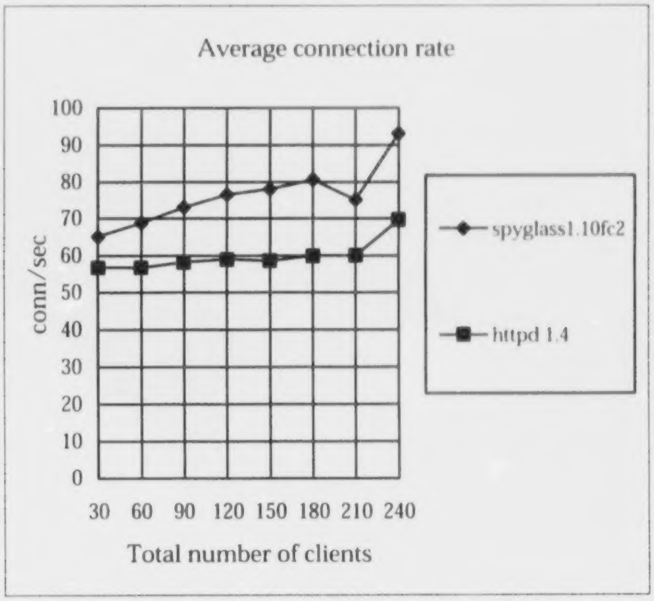


Figure 3: Average connection rate on single-processor configuration (conn/sec)

The results on dual-processor configuration are in [Tab. 3]. It shows the opposite of [Tab.2]. HTTPd of multi-process model is better than Spyglass. HTTPd handle 16.64 connections more than Spyglass, and the average latency time is lower by 0.28 second. In addition, it transfers more data by 0.99 Mbit per second. The average connection rate is charted in [Fig. 4]. The peak difference takes place at 240 clients, and HTTPd is 50 percent better than Spyglass

Total number of clients	Average connection rate (conn/sec)		Average latency (sec)		Throughput average for all connections (Mbit/sec)	
	HTTPd 1.4	Spyglass 1.10fc2	HTTPd 1.4	Spyglass 1.10fc2	HTTPd 1.4	Spyglass 1.10fc2
30	94.83	72.03	0.3175	0.4247	5.12	3.85
60	92.03	74.70	0.6640	0.8433	5.00	3.99
90	94.87	80.62	0.9727	1.1602	4.94	4.22
120	95.55	83.92	1.3374	1.5259	5.15	4.56
150	97.45	85.20	1.6709	1.9779	4.87	4.46
180	94.58	87.32	2.0703	2.2499	5.63	4.64
210	94.93	86.82	2.4328	2.6209	5.26	4.09
240	117.35	77.82	2.1268	3.0452	6.23	3.82

Table 3: Results on dual-processor configuration

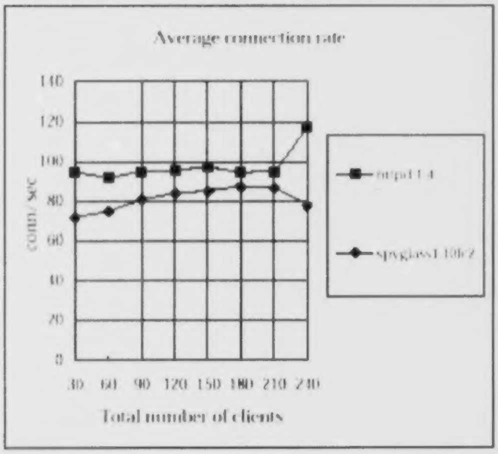


Figure 4: Average connection rate on dual-processor configuration (conn/sec)

Conclusion

In this paper, we attempted to analyze two web server process models. The results show that the single-process model has higher performance in a single-processor system. We believe that it is because it does not fork a new process for each connection. But, this model cannot utilize the resources of multi-processors system. In

BEST COPY AVAILABLE

other words, multi-processing capability of multi-processor systems is not exploited well. On the other hand, the multi-process model achieves higher performance in a multi-processor system because several processes can service each client simultaneously. However, in a single-processor system, it produces additional overhead for making a new process and context switching between processes. We believe our results are not complete and we plan to investigate the effect of process model on web server performance with more number of processors.

References

[NCSA 1996] NCSA HTTPd Development Team, "NCSA HTTPd Overview",
[URL:http://hoohoo.ncsa.uiuc.edu/docs/Overview.html].

[SpyGlass 1996] SpyGlass, Inc, ``About the Spyglass Web Server Benchmarks",
[URL:http://www.spyglass.com/products/server_results.html].

[Gene and Mark 1995] Gene Trent and Mark Sake, WebSTONE: The First Generation in HTTP Server Benchmarking,
[URL:http://www.sgi.com/Products/WebFORCE/WebStone/].

[Michael 1996a] Michael Blakeley, ``SGI WebFORCE Challenge S vs. Sun Netra i20",
[URL:http://www.sgi.com/Products/WebFORCE/WebStone/sun-ss20/sun-ss20.html].

[Michael 1996b] Michael Blakeley, ``SGI WebFORCE Challenge S vs. Pentium 100MHz Windows NT",
[URL:http://www.sgi.com/Products/WebFORCE/WebStone/nt-p100/nt-p100.html].

Infusing Web Content with Educationally Relevant Indices

Rose Kotwas, Lincoln High School
Janet Luongo, The Discovery Museum
Miriam Masullo, IBM T.J. Watson Research Center
Joan Musgrave, IBM T.J. Watson Research Center
Ben Ortiz, The Discovery Museum
Ray Ramirez, North Central Regional Educational Laboratory
Joanne Tursky, IBM T.J. Watson Research Center

Webnet'97
Toronto, Canada
November 1997

Extended Abstract

The sudden online availability of content on the World Wide Web (WWW or Web), presents great educational potential. Images, video and audio from a endless variety of sources can be used to enhance lesson plans, student projects, and ultimately curriculum. But, with increased focus on educational standards and accountability, it is necessary to separate content applicability to goal oriented education from content that becomes a distraction from educational goals. This differentiation is a subjective and difficult task that must be systematized if teachers are expected to cope with the deluge of material on the Web.

EduPort (1) has provided a methodology for indexing content with respect to educational goals and standards that can be implemented as an added index catalog to digital collections (libraries and museums). The approach is to provide an added layer of metadata to any digital content collection that reflects the educational value of the content.

The guidelines used to create EduPort metaindices were based on work done in support of educational standards. The indexing scheme used includes the fundamental information needed in order to develop curriculum. Once that information is systematically acquired and cataloged, curriculum blocks can be defined that target educational goals and standards; and lesson plans can be designed that deploy useful content and map to a curriculum framework. Most importantly, the "digital curriculum library" that is created by this process, can then be shared as a digital collection, in its own class, of teaching exemplars.

To illustrate the potential of the approach, two digital collections were indexed for EduPort after their initial creation as Web pages. In the domain of Art, a private collection, or online digital gallery, was indexed by the artist with respect to EduPort metadata(2), and the art collection of a museum is currently being indexed in this fashion by the museum curator. The educational perspective on the content was then enhanced with comments from two art teachers. This kind of metadata can be incorporated into any other media collection, where there is a desire to make them more useful for education.

The resulting home pages, in this case, were thus enhanced with educationally relevant data, making the content more useful to teachers in search for online sources to enhance curriculum. The next aspect of the approach is to collect the resulting content applications in the form of lesson plans (how teachers are using content so indexed), and to formulate general purpose curriculum blocks from these lesson plans. These curriculum blocks are stored back into the EduPort Digital Curriculum Library for reuse by other teachers, or as "design illustrations" for creating new lesson plans.

Such process begins to systematize the task of making the Web useful for education, in an open-ended, collective approach that involves the teachers as well as the content providers. The collective global resource is enriched by every single object on the Web that is so indexed, and by every lesson plan that is created with these objects.

The task of making sense out of what is on the Web today befalls the teacher. Search tools, and automatic content monitors fall short of turning content "found" into content "used" to create teaching material. A worse problem, not generated but perhaps aggravated by the new Web environment, is that content must be found over and over again, by every teacher that needs to use it. The results of "educationally oriented searches" are not systematically kept and shared. Therefore a teacher in Connecticut must repeat the tasks that a teacher in Nebraska performed a while ago, and sadly these teachers do not connect in systematics ways to share their experiences.

The emergence of national standards provides a point of commonality for sharing teaching experiences, and for indexing content in a generally accepted and useful form. Like standards, indexing of content and sharing of teaching exemplars requires consensus. It is unlikely that one single perspective for extracting value from free flowing content on the Web will ever be found, nor is that desirable. But, some initiatives to add educational value to content can be adopted, and that is desirable, as a practical approach to teacher productivity in a climate of wild content growth.

A National Digital Library of Educational Media is needed as the layer that separates all things on the Web from educationally useful media on the Web. Not only must media be educationally useful, but applicable to curriculum, for teachers to be able to cope with the vastness of the digital medium. This presentation demonstrates what such consensus would accomplish and how it can be reached.

General Reference

(1) http://ianrwww.unl.edu/eduport/eduport.htm

(2) http://www.research.ibm.com/people/m/musgrave

Towards a Web Operating System (WOS)

Peter Kropf,[1] John Plaice[2] and Herwig Unger[3]
[1] Université Laval, Département d'Informatique, Sainte-Foy (Québec), Canada, G1K 7P4
[2] University of New South Wales, School of Computer Sc. and Eng., Sydney 2052, Australia
[3] Universität Rostock, Fachbereich Informatik, D-18051 Rostock, Germany

Abstract. We discuss the concept of computing on the Internet or the Web based on a demand-driven technique called eduction. Because is is impossible to define fixed sets of operating system functions for all services and to keep a complete catalogue of all available resources, incrementally built versions of a Web Operating System (WOS^{TM}) are proposed to offer an exhaustive variety of services available on a network. The WOS basically consists in a collection of eductive engines and warehouses all available from the Web.

Introduction

With the rapid development of new forms and concepts of networked and mobile computing, it is increasingly clear that operating systems must evolve so that all machines in a given network, even the Internet, can appear to be controlled by the same operating system. As a result, the world-wide interconnected networks, commonly called the Internet or the Web, could potentially be supported and managed by a giant virtual operating system [Reynolds 1996].

For example, initially the World Wide Web was created to allow one to view remote hypertext pages on one's own machine, thereby facilitating collective work among geographically removed collaborators. Soon after, virtual pages, generated on the fly using tools such as cgi-bin [NSCA-CGI 1996], allowed the widespread remote execution of programs. More recently, with languages such as Java [Arnold and Gosling 1996], it has become possible to download fully executable programs to one's own machine, and then to make them run on that machine. However, there is no general means for taking an arbitrary program and having it run somewhere on the network.

There are several reasons that this last possibility is actually essential. First, with the development of network-centric computing, there will be more and more limited-capacity machines (slower processors, limited memory or storage space, etc.), such as the NC computers [Sun-NC 1997], that will be forced to use more powerful computers on the network to effect any non-trivial tasks. Second, an arbitrary program on the network might just be incapable of running on the local machine, simply because it is the wrong platform (hardware, local operating system, running applications, etc.)

Implicit in the above discussion is the heterogeneous nature of most networks. The transparent use of such heterogeneous networks of computers has been partially addressed in work on metacomputing, whose objectives are to transform a network into one single computer system [NCSA-Meta 1997]. Recent developments in operating systems such as Inferno [Lucent 1997] or JavaOS [Sun-JavaOS 1997] provide the user ubiquitous access to resources and information. However, the Web or the future global information infrastructure is more than just a metacomputer or a networked system of computers seen as a virtual machine run by a virtual (network) operating system, in that there is no complete catalog of all resources available. Moreover, such a catalog is infeasible, because of the highly dynamic and distributed nature of the Web or the Internet, continually integrating rapidly developing technologies.

Web Operating System: WOS

As a result, any attempt to design one single operating system offering a fixed set of resource-management functions will have difficulty adapting to technological innovation or to new demands. Therefore, there is, such as proposed in [Ben Lamine et al. 1997], a need for a Web Operating System (WOS), which would make available, to all sites on a network, the resources available on that network, or at least a reasonable subset thereof, to effect computations for which local resources are missing. These resources could be of many forms, including processor speed, available memory or storage space, available operating systems or applications, and so on. In order to deal with the dynamic changes in the system, the Web Operating

System should be a versioned system, in which different versions of the operating system are running simultaneously on the network.

What distinguishes the Internet from classical distributed systems is the fact that there is no complete catalog of all resources available and central decisions making for resource allocation is not acceptable or even impossible. Rather, the Web Operating System (WOS) [Ben Lamine et al. 1997] should be a versioned system, in which different versions not capable of dealing with a particular request for service, then pass it on to another version, as currently done for packet routing. Generalized software configuration techniques, based on a demand driven technique called eduction [Plaice et al. 1997] are being developed, that can be used to define versions of a WOS to be built in an incremental manner. Software and hardware (description) repositories or warehouses will provide the necessary components for fulfilling a service requested. The kernel of a WOS would be a general eductive engine responding to requests from users or other eductive engines and fulfill these requests using its warehouses.

The WOS would then work in the following manner. A request would be placed by a user to run a particular program or to initiate some service. The programs or services might be located at different sites of the network. The eductive engine would then decide whether it is capable of dealing with the request or whether it will pass it over to some other eductive engine until finally one engine accepts responsibility for the request. Once all the resources (programs, services, hardware) become available, then the program could be run and the requested service fulfilled.

Conclusions

The general aim of our approach is to develop a family of services for illustrating and studying the concept of a Web Operating System (WOS) based on one single underlying concept, the demand-driven computation using simple warehouses, which hold and provide all the necessary information a system may offer to a request. The WOS will thus take the form of an add-on service to existing operating systems, such as UNIX or NT. The ongoing work includes (1) production of sample resource managers and warehouses, together with the necessary automatic broadcast or 'resource-mining' mechanisms, (2) the implementation of a sample series of WOS-services (e.g. typesetting services, graphics processing, interactive simulations, etc.) and (3) implementation of a prototype user-interface based on browser-like forms to specify user (application) requests, which includes new 'data-mining' search engines.

References

[Arnold and Gosling 1996] K. Arnold and J. Gosling. *The Java Programming Language*. Addison Wesley, Reading, MA, 1996.

[Ben Lamine et al. 1997] S. Ben Lamine, J. Plaice, and P. Kropf. Problems of computing on the WEB. In A. Tentner, editor, *High performance computing 1997*, pages 296–301, Atlanta, Ga., SCS,1997.

[Lucent 1997] Inferno. Lucent Technologies, `http://www.lucent-inferno.com`, 25 July 1997.

[NCSA-CGI 1996] The Common Gateway Interface. National Center for Supercomputing Applications, `http://hoohoo.ncsa.uiuc.edu/cgi`, 5 February 1996.

[NCSA-Meta 1997] NCSA Metacomputer. National Center for Supercomputing Applications, `http://www.ncsa.uiuc.edu/General/CC/CCMetacomputer.html`, 25 July 1997.

[Plaice et al. 1997] J. Plaice and S. Ben Lamine. *Intensional Programming II*, chapter Eduction: A general model for computing. World Scientific, Singapore, 1997. (to appear)

[Reynolds 1996] F. Reynolds. Evolving an operating system for the Web. *IEEE Computer*, 29(9):90–92, 1996.

[Sun-JavaOS 1997] JavaOS. Sun Microsystems, `http://java.sun.com/products/javaos`,25 July 1997.

[SUN-NC 1997] Network computer reference profile. Sun Microsystems, `http://www.sun.com/smi/Press/sunflash/mncrs-profile.html`, 25 July 1997.

An User Authentication Method for WWW usingKey Distribution Algorithm and Random Sequence

Yun-ho LEE
Korea Telecom R&D Group, Republic of KOREA
leeyh@time.kotel.co.kr

Young-sun KIM
Korea Telecom R&D Group, Republic of KOREA
yskim@time.kotel.co.kr

1. User Authentication Problem over WWW

So far, username/password scheme, the most widely used method for user authentication in WWW, has a serious weakness from a security point of view, as it sends username and password plainly over the network. In cryptography, a variety of authentication algorithms other than username/password studied to increase the safety. However, most algorithms were based on the connection oriented environment and hence they can not be applied to WWW directly. Therefore it is necessary for new approaches to apply modern cryptographic user authentication algorithm to WWW.

2. Key Distribution and Random Sequence

In this paper, in order to overcome the connectionless property of WWW, we propose a method of information security using RSQ (Random SeQuence) to check user authenticity every time client requests connection to the server. If user is authentic, he can generate a series of RSQ using his own secret information. In this case, we have to consider a way to share common seed value between client and server. This sharing problem can be solved by using KD (Key Distribution) algorithm in cryptography. Both client and server can share a common key created after key distribution as the seed value. In addition, as the generation of RSQ can be dealt with high speed by using the conventional cryptosystem, the overhead of communication between client and server will be trivial. Furthermore, using KD algorithm, the only verified user and server can share common key and nobody can know the common key. Thus server can verify user's authenticity by compare the two RSQs, one is generated by client and the other is generated by server, which should be the same.

3. Propose An User Authentication Method

Phase 0. (Server Preparation)

0-1. The server choose prime number P, a(primitive element in mod P), CCS(Conventional CryptoSystem) as public information, and K_s ($1 < K_s < P$) as secret information.

Phase 1. (Pre-register)

1-1. Client A register : choose public ID_a and secret K_a ($1 < K_a < P$)

1-2. Server stores client A's information, ID_a, K_a and RSQ(=0 initially).

Phase 2-a. (Request Connection - DH-like KD algorithm - Interactive version)

2-1. User A requests connection and sends ID_a

2-2. Server chooses random number R_s ($0 < R_s < P$), calculates $I = a^{(KsRs)} \bmod P$, and sends a, I, P and necessary functions to user A.

2-3. Client A processes the follows using his secret key K_a;

- chooses random number R_a ($0 < R_a < P$)
- calculates $V = a^{(KaRa)} \bmod P$, and the first RSQ value $R_a = I^{(KaRa)} \bmod P$
- calculates the second RSQ value $R_a' = CCS(R_a, K_a)$ and sends ID_a, V, and R_a' to server.

2-4. Server calculates $R_s = V^{(K_sR_s)}$ mod P and $R_s' = CCS(R_s, K_a)$.

2-5. If $R_s' = R_a'$ then server recognise the client as valid.

Phase 2-b. (Request Connection - DH-like KD algorithm - Non-Interactive version)

2-1. Client A requests connection and processes the follows;

- chooses random number R_a ($0 < R_a < P$)
- calculates $V = a^{R_a}$ mod P
- calculates the first RSQ value $R_a = V^{K_a}$ mod P and the next RSQ value $R_a' = CCS(R_a, K_a)$ mod P
- sends ID_a, V, and R_a' to server.

2-2. Server calculates the first RSQ value $R_s = V^{K_a}$ mod P and the next RSQ value $R_s' = CCS(R_s, K_a)$.

2-3. If $R_s' = R_a'$ then server recognise the user as valid.

Phase 3. (Communication - checking RSQs using CCS)

3-1. When client A requests information from server, client A calculates the next RSQ value $R_a = CCS(R_a, K_a)$, sends it to server with ID_a.

3-2. Server verifies the users' validity as follows;

- calculates the client A's next RSQ value $R[A] = CCS(R[A], K_a)$. R[A] means client A's current RSQ stored in server's table or database.
- if $R[A] = K_a$, then server provides services requested, otherwise, disconnects connection.

3-3. If server does not detects any connection request from client A or user selects disconnection, server clears the value of R_a to zero.

4. Simulation

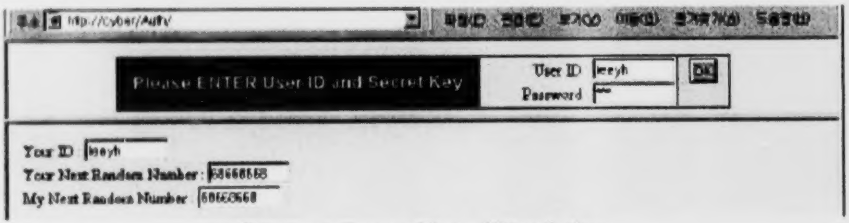
The user authentication method presented in this paper, checks if a user can generate correct RSQ(Random SeQuence) values every time he connects to server to overcome connectionless environment. In order to verify our method presented in this paper, we simulated our method by implementing it.

Hardware		Software	
CPU	Pentium Pro 200MHz x 2	O.S	WindowsNT Server 4 with Service Pack 3
RAM	64MB		
HDD	4GB	Web Server	MS IIS 3.0 + ASP
		Script	VBScript
		Browser	MS IE 4.0 preview 2

Table 1: Simulation Environment

We implement the 2b method, non-interactive version. First, we divide screen into 2 frames -main frame and authentication frame- using frame facility in HTML. User enters his ID and Key in the authentication frame, receives/sends information from/to server in main frame.

We select values for a and P as 7, 997 repectively. However, they are insufficient to be applied to real application as the value of P is too small. It is recommended that the value of P must be a prime number greater than 2^{64}.



User ID leeyh
Password

Your ID leeyh
Your Next Random Number : 68668668
My Next Random Number 68663668

Figure 1: Screen Shot of Simulation

Here is the detail information request/retrieval process between user and server.

a. Client loads a HTML page.

b. Client reads current RSQ value from authentication frame by loading time automatically.

BEST COPY AVAILABLE

c. User selects information needed.

d. Before send form data to server, client calculates the next RSQ value and updates it to authentication frame and stores it in the form field.

e. Server updates A's RSQ value stored in table (or database) if he receives correct ID and RSQ

f. Server provides information requested.

References

[Berners] http://www.w3.org/pub/WWW/People/ Berners-Lee/

[Diffie et al. 76] Diffie, Hellman (1976), "New Directions in Cryptography", *IEEE Trans. On Information Theory*, IT-22

[Lee et al. 92] YH Lee et.al. (1992), "A Study on the Identification and Key Distribution Protocol Using ID-Based Zero-Knowledge Interactive Protocol", *Journal of The Korean Institute of Information Security and Cryptology*, Vol. 2, No. 1

Using Web Technologies to Support a Knowledge Management Culture

Paul Lefrere
Institute of Educational Technology
Open University
UK
P.Lefrere@open.ac.uk

Andy Stone
Institute of Educational Technology
Open University
UK
Andy.Stone@open.ac.uk

INTRODUCTION: web technologies and organisational knowledge

From a technological point of view, there is now a wide choice of candidate technologies for providing ready access to organisational information. Many of these are web-based or can be integrated with web-based systems. For example, it is now relatively straightforward to provide everyone in an organisation with access to a legacy database or a data warehouse via an intranet or extranet. Likewise, it is relatively easy to provide personalised "power pages" via an intranet, as well as personalised news channels and targeted email via "push" technologies such as CastaNet's Marimba.

Information is not a free good, in economic terms, since its assimilation and utilisation requires an appropriate level of understanding. Information is only of value within a context where other forms of knowledge are brought to bear. Indeed, for a number of reasons it may have negative attributes. There is much current concern with 'information overload' - that is, too much information swamping an individual or an organisation's ability to assimilate and use it. The following quote from Herbert Simon illustrates this: "*In a world where attention is a major scarce resource, information may be an expensive luxury, for it may turn our attention from what is important to what is unimportant.*" (Simon, 1978). What is important is a matter of judgement, and depends on other forms of knowledge being brought to bear, including "*meta-knowledge*" (knowledge about knowledge).

What knowledge is important?

For individuals, important kinds of knowledge include:

- knowledge which is up to date or can be refreshed (as in professional updating courses, increasingly delivered at point of need, via the web)

- knowledge which is lasting (as in courses on how to learn; courses on how to identify and forget knowledge that is no longer useful, and courses on how to identify and exploit unseen value in one's existing knowledge).

Organisations have similar needs, but increasingly have to operate in a context of increased rates of change, competition and market turbulence. They need new ways to compete effectively.

One important process, which can be supported using web technologies, is to recognise or rediscover assets which an individual or an organisation already have, but are not being used to their full potential. These include procedural knowledge, patents, copyright, brands, R&D, licensing opportunities, innovative use of assets such as databases, and so on. These provide opportunities to innovate, cut costs, save design time, reduce time-to-market, etc.

Knowledge is different from information

It is generally accepted that information technology has brought about a qualitative change in our society, by making it easy to produce, reproduce and communicate vast amounts of data and information electronically. In this context, 'more' information is not necessarily 'better'. Unlike material commodities, in the economics of information, 'more' is worthless unless it is:

a) *different*: is new to a given user, rather than replicating or confirming existing information), and/or
b) *usable*: in the sense that someone who wishes to use an information source can access appropriate information, understand it and utilise it within a relevant context and time frame, and at an affordable cost.

Charles Handy has recently claimed that the future lies in a 'three-i' economy, with organisations adding value through the application of information, ideas and intelligence (Handy, 1995). For us, as we shall argue, the effective deployment of web technologies requires consideration of pertinent human factors (social, cultural, individual), within more powerful frameworks which provide new insights into the nature and context of organisations and their information needs. One such framework is that of "Knowledge Management". The advent of such technologies has highlighted a number of problems, such as information overload, which cannot be solved by consideration of technological factors alone.

The claim, therefore, for the emerging inter-discipline of knowledge management, is that knowledge must be the focus for analysis, and that organizations must find ways in which to manage the processes by which knowledge is created and applied.

What is Knowledge Management (KM)? How can web technologies support it?

For the purposes of this paper, we define Knowledge Management (KM) as "the process of continually managing knowledge of all kinds to meet existing and emerging needs, to identify and exploit existing and acquired knowledge assets and to develop new opportunities". Web technologies can support KM by:

- supplementing internal knowledge-sharing activities such as "knowledge fairs"; by using a KM intranet to provide transparent access to a wide range of heterogeneous internal information sources, including legacy databases, evaluation reports and discussion spaces

- using a KM extranet to support knowledge sharing by trusted third parties (for example, the European SOCRATES-funded Student Virtual Mobility project).

- providing facilities for users to annotate web pages on the intranet and extranet (we are developing such a facility as part of an internal project at the Open University)

People can contribute know-how using the web in different ways. These can be described as "passive" (publishing static HTML), which can be picked up by search engine "spiders" (as used by Alta Vista etc). Alternatively, they can take a more "active" approach (post to an email list/newsgroup), email lists can be searchable, and newsgroups are via technologies such as Deja News).

The issue here is not so much HOW information is submitted, but how meaningful knowledge is constructed from this information. This is where web technologies come into play:

Spiders are useful, but are retrospective, reactive applications. We feel that it is necessary to move towards proactive applications (where as soon as information is entered, it can be accessed as knowledge, at the right time in the right way with the right perspective - i.e. with appropriate delivery method for that particular user)

Such applications should include the adoption of a combination of the following:

- push technology

to enable proactive delivery of relevant information in real-time (or, to be read at a time which is appropriate for you)

- X500/LDAP support

directory services can be used to enable the acquisition of contextual information; this facility is featured in Netscape Communicator 4.0 and previously there have been web-X500 gateways, but without implementation with a coherent KM strategy, the benefits will be suboptimal.

Conclusions

Using web technologies effectively in an organisational context does not mean gathering together every piece of information in the organisation and then providing access to it via an intranet. A lot of that information is likely to be useless, irrelevant, outdated or too costly for individuals or their organisations, or reflects the past that we're trying to escape.

A better approach, which can be informed by KM insights, is to seek out or create specific kinds of knowledge (some of which an organisation may not even know it has) for specific purposes (such as competitive advantage or greater efficiency). Those kinds of knowledge include knowledge which has long been known but not applied to the current problem. So the issues of uncertainty and complexity have a particular importance here - how do we know we have useful knowledge? How do we know that our successes are due to its exploitation?

These conceptual challenges do not mean that meaningful action cannot be taken. Intelligent use of KM approaches can suggest an agenda for the development of action-oriented goals for managers, organisations and networks of organisations, including:

- the formulation and implementation of strategies for developing, acquiring and applying knowledge and deploying it via the most appropriate technologies, such as web-based technologies and mobile telephony;
- the daily improvement of the business processes in an organisation, with a focus on knowledge development and use; and
- the monitoring and evaluation of knowledge assets and their effective management.

References:

[Simon, 1978]
H Simon, Traditionality as process and product of thought, American Economic Review 68, (2), 1-16 (1978)

[Handy, 1995]
C Handy, Trust and the virtual organisation, Harvard Business Review, May-June, 40-50 (1995)

NDSU's Tailor-made Web site Revamp

J. Ludvik Herrera Lopez, Webmaster
Information Technology Services
North Dakota State University
United States of America
herrera@prairie.nodak.edu
http://www.ndsu.nodak.edu/

Introduction

North Dakota State University (NDSU) in Fargo, ND, offers some 80 majors in 200 specialties in Science and Mathematics, Engineering and Architecture, Agriculture, Human Development and Education, Business Administration, Pharmacy, Humanities and Social Sciences, and University Studies.

All of these areas attract great diversity to the university and its website. We have recently redesigned the website so users can find and authors can create information quickly and easily. The new design accomplishes five objectives: comprehensive site design and management; consistent page layout; effective marketing of the university's services; ease of navigation; and eye-catching graphics.

We have succeeded by building a *NDSU Web Team* whose members represent different points of view and various backgrounds. The team is led by the University Webmaster and the Director of University News and Publications. Assisting them are graphic designers, a multimedia technologist, an instructional designer, a UNIX consultant, and the Director of Learning Technologies. At the beginning of the redesign process, the only practical thing the members of the team had in common was *HTML* encoding experience and a desire to achieve the five objectives.

Comprehensive Site Design and Management

NDSU's website grew extensively over the past two years. Statistics on campus development and use showed that faculty, staff and students had begun to use *the Web* as a primary information resource. The information on our website, however, was becoming out of date, the complexity and the volume of information were making individual parts of the site disorganized.

As a first step toward redesign, the Webmaster and Director of University News and Publications drew a site-map of the entire site. This map gave them a view of all the up-to-date information on the site -they were able identify to whom it belonged, who was responsible for its maintenance, and links to and from each unit's or department's pages. The site-map is now the "*backbone*" to which the team attaches new areas or sections of information.

The Webmaster, the Multimedia Technologist, and the Instructional Designer next defined a flow-charting or storyboarding scheme that those proposing new construction could use to communicate their plans to those who would review, approve and execute the work.

Consistent Page Layout

A consistent look, feel and function to our website is important. It reminds users where they are as they browse our site and that the information they are retrieving is reliable. The NDSU Web Team constructed different templates for different levels of information for campus developers use. Training faculty, staff and students is one key to successfully implementing consistent page layout. Guidelines and a style book, which strive to balance uniformity and individuality in web page design, help inform the campus community about the university's intentions to use *the Web* as a marketing tool.

Marketing of the University's Services

As a marketing tool, the NDSU Website uses "advertisements" on its homepage. Banners are randomly displayed in the center of the main page. When users "hit" or reload the homepage, for example, a new banner appears.

We use these banners to help orient new users to the organization of our site, to alert users to timely information about campus and events, and to direct users' attention to new information or activities in various departments on campus. We add or change the banners as new information is added to the site or to make announcements. The site continues to have a consistent look and feel so users always recognize the site, but the site is also more dynamic. Site portions are featured on the main page and these features are changed regularly.

Ease of Navigation

We have created several versions of the main page suitable for various browsers. The Webmaster and the UNIX consultant determined how and what information to gather from the client browser in order to customize the delivery of the main page to users. The UNIX consultant wrote a CGI script which gathers the necessary information about the client browser's capabilities, and returns the page best suited to the browser and the speed of the client's network connection.

There are three versions of the main page: a JavaScript version with animated "advertising" banners and interactive graphics, a low-bandwidth, non-JavaScript version, and a version requiring the Macromedia Shockwave plugin. This "shocked" version incorporates sound and animation in a stand-alone, navigation window. Navigation selections made from this window are returned in a separate browser window. That way, the features of the main page of our site are always available, no matter how deep the user traffics in our site.

Eye Catching Graphics

The Web Team is trying to encourage the NDSU campus to use graphics created by professionals so the site has integrity and portrays the campus in a professional and competitive manner. As an inducement, the Webmaster has created an interactive form on the web which allows web developers to choose the Official NDSU headers, subheaders and buttons for their pages. The Webmaster's development webserver renders the selected image and its HTML code so that the web author can copy the image location or insert the tag in their new document. These graphics are optimized for download time. The graphics are also built using web-safe palettes, all saved in two standardized sizes, with preset color depths, as gifs and all the official graphics use NDSU original designs, copyrighted or trademarked logos.

Intelligent Indexation of Compound Document using Images

Nicolas Malandain, Mauro Gaio, Maurice Szmurlo.
GREYC - CNRS UPRESA 6072
University of Caen, Esplanade de la Paix, 14032 Caen Cedex, France
{Nicolas.Malandain, Mauro.Gaio, Maurice.Szmurlo}@info.unicaen.fr

Introduction

Efficient software for the general public, that allows merging different media, changes the way documents are composed. The best example is the Web where documents contain more and more images. Unfortunately indexing engines don't use the information contained in these images.

There are currently two strategies for image indexation. The first one is based on global shape, color and texture analysis. Current state of the art allows to operate just on a set of non-complex images. For instance, problems are still not solved when the represented objects have a quite similar shape. Assume three images representing the Eiffel Tower, the Ariane rocket and a chicory : none can be distinguished from each other only by their shape. Even, color analysis does not help much in the differentiation of the two last images. This method is therefore mainly of interest on highly heterogeneous image databases: a query-by-example based on an image only retrieves globally similar images without taking into consideration the semantics of the content.

The second strategy [Neil 1996] attempts to overcome this drawback by taking into account words in the title and associating them with the shapes detected in the illustrations. A request may be submitted either by an image or by keywords. In the first case, the previous strategy is used. In the second case, the images associated with the keywords are retrieved, and the associated shapes are used to proceed a research as in the first case.

This position paper presents the current state of our analysis of a tool which improve automatic indexing for compound document. This indexing tool add to classical IR engines the information extracted from a basic semantic analysis of the links between text and geographical illustrations and some expressive feature contained in these illustrations.

A geographical illustration allows the reader to see the spatial organization of the phenomenon evoked in the text. Actually to test the feasibility, only statistical maps[2] are taking into account.

Searching "where it's about" the Map Content

Our work hypothesis is that in almost all scientific or technical compound documents, text and images had powerful semantic links realizing a global coherent meaning. As a practical example we have based our corpus[1] on geographical illustrations as thematic maps or statistical charts. We have chosen maps because they are always described by a comment in the running text and not only by a simple caption. Note that geographical illustrations may moreover contain information that is not explicitly expressed in the corresponding running text.

In the indexing process of a document from a paper source, the document is rebuilt in an electronic format but no link is made between text and images. Two types of links exist, on the one hand there are explicit links created by a reference (e.g. "see fig 1"), on the other hand there are implicit links made by the author when he evokes an image (in our context a map) in the text without citing it explicitly. To create an implicit link, the author uses a vocabulary shared by the title of the map and the analysis described in the text.

To extract information from both text and maps, we need to find these implicit links. We use a coarse filter based on Information Retrieval techniques (IR such as the statistical method $t_f * i_{df}$) [Salton 1988]. Its goal is to search the common vocabulary int the running text. It produces a sorted list of paragraphs according to the weight of their link with the map. Note that we must adapt the IR algorithm according to the data of our context. Firstly, documents of a common retrieval system are replaced by the paragraphs of the text (in which we search the information). Secondly, we necessarily know that somewhere the map is evoked in the text.

Afterwards, we will extract specific information from a subset of highly scored paragraphs and their surrounding (e.g. the containing section).

Geographical Information Extraction

The first step of this process consists in building what we call the *dynamic knowledge* by pointing out from the retained paragraphs some geographical semantic information that will drive the analysis of the map. As the author phrases some specific areas because of their importance in his geographical, the first useful clues are names of towns, regions or any other named geographic objects. The second clues are expressions about spatial directions or locations that are searched for (e.g. "... in the north of Paris ..." or "... around the Normandy...").

Classical rules for proper names detection will be used and ad hoc patterns will match the spatial information (e.g. in the north [some place]). In our current experimentation, we need spatial information to identify and locate the geographical objects extracted. These informations are stored in a Geographical Information System (GIS) for perform this task, such informations are named *static knowledge*.

Map Basic Features Extraction

This second step consists in map characterization. Firstly geographical localization of the map contents is estimated. The "*dynamic knowledge*" cuts down the different possible locations, then with the outlines of the map and the geometric informations stored in the GIS the map is geographically identify. Secondly the real contents analysis is processed. It consists in searching some basic features like groups, boundaries between areas (north, south)... Clustering methods have been used to perform this task, but to avoid some errors, the clustering process is driven by the "*dynamic knowledge*" previously extracted.

Conclusion

We are currently working on the implementation of such a tool. The first goal was to define the global architecture of the system, to adapt IR algorithm and to produce rules to build the *dynamic knowledge*. Our purpose is now to extract information from the map using *dynamic knowledge*. Currently we only want to index the description of the map with the document In the future, for a driven semantic interpretation of more specific feature of the map we plan to add to the *Static knowledge* the geographers common knowledge.

References

[Neil 1996] Neil. C.Rowe (1996). *Precise and Efficient Access to Captioned Picture Libraries: The MARIE Project*. http://www.cs.nps.navy.mil/research/marie. Department of Computer Science. Naval Postgraduate School.

[Sargur et al. 1994] Sargur, N.Srihari, Stephen, W.Lam, Jonathan, J.Hull, Rohini, K.Srihari, and Venugopal Govindaraju. Intelligent Data Retrieval From Raster Images of Documents, 1994,*Digital Libraries'94 conference* .

[Salton 1988] Salton, G. (1988). *Automatic Text Processing*, Addiso-Wesley (pages 229-374)..

---

[1] atlas containing documents composed of statiscal maps and charts, and text.

[2] maps of country divided in regions containing a color that corresponds to a value.

The *Virtual Moderator*™ Project

Jacques Malouin, Faculty of Education, University of Sherbrooke, Québec, Canada
jmalouin@courrier.usherb.ca

As suggested by a growing number of specialists, collaborative work, which is now made possible by information technology, is perhaps the most efficient way to learn and certainly to get accustomed to constructive critics between peers. The development of a new collaborative tool that we have called *Virtual Moderator*™ has involved technical and programming challenges, but this short paper is centered on pedagogical aspects of the design. We first present the context, the goals and the rational for developing this tool. The system itself is then briefly described and some perspectives are outlined as a conclusion.

Development Context

Although the *Virtual Moderator*™ system can be used in numerous and various situations where members of a group have to collaborate in a task even if they are not physically together, it was first designed as an educational application to be used by students working collaboratively through the superhighway. Those students were using Internet in two ways: asynchronously, they shared texts, pictures et graphs and use E-mail system to communicate with each other; synchronously, they « met » (usually in groups of 5 to 6) at a specific time to experiment on-line collaborative works.

In the case of asynchronous tools, it was relatively easy to find applications that satisfy our needs. But with respect to synchronous tools, the situation was quite different. Every application we experienced or for which we got documentation had to be eliminated for some of the following reasons :

- They were part of a broader system and consequently, were too complex and too heavy systems.
- They were limited to one platform only at a time and consequently, lack universality of application.
- They involve real time pictures or/and sounds and consequently, were far from a minimum level of performance especially when modems had to be used.
- They had poor design interface and were not taking into account recent developments in the field of cognitive psychology, especially with respect to concepts like distributed knowledge, cognitive intelligence, etc.
- Above all, they did not involve a sub-system that would regulate the interventions of the various participants and facilitate orderly and efficient collaborative works.

Development goals

Because we were not able to use or adapt an appropriate system amongst existing applications, we decided to built a new one that would reflect new developments in the field of cognitive psychology and especially social and contextual factors that were incorporated in the new learning models. The system to be developed should also make collaboration easier and, in particular, allow two or more students to plan synchronously and interactively a common solution to a problem and to share the cognitive load required for the accomplishment of the task.

Essentially, *Virtual Moderator*™ system was designed with the idea of

1. simulating the context of round table discussion where a group of people collaborate at a common task through various kind of interventions, proposals and votes ;

2. simulating the presence of a human moderator who would take note of the requests for an intervention, indicate when each turn comes up and, in general, facilitate collaboration among members of a group working at a common task while being in different locations.

The *Virtual Moderator*™ interface has been carefully designed to provide all the tools required for a fruitful collaboration while avoiding unnecessary and disturbing objects and gadgets which may be technically interesting but which may interfere negatively with the cognitive process. We think that such an environment will make collaboration easier because participants will concentrate on the task to be accomplished.

Characteristics of the *Virtual Moderator*™ system

The *Virtual Moderator*™ design is unique in two ways. First it is unique by its design interface and the choice and arrangement of functions involved in the system. Second, it is unique by its built-in sub-system of regulation of interventions which is described here under.

Design Interface

In the *Virtual Moderator*™, every tool and object of the interface has been designed with the round table concept and the human information system in mind :

- the photography and name of each member of the group which appear as he or she joins the work session ;
- the chat window, associated with the working memory, which holds current individual interventions as they take place ;
- the white board window, associated with the long term memory, which holds shared constructed objects of the group ;
- the previous sessions reports window, which is also associated with the long term memory ;
- the consultation windows which allow each participant to express his view on specific questions and to monitor other participants' position on those questions ;
- the private message window which allows each participant to *whisper* a private message to another participant during the meeting.

Virtual regulation of interventions

The moderator sub-system has been designed to regulate all the interventions of the participants as if a human moderator was conducting the meeting with smoothness and efficiency. This sub-system incorporates the following objects and tools :

- a *hand raised* icon which allows any participant to indicate that he wants to intervene (or that he does not want to intervene anymore) ;
- colored and numbered frames which identify the participants who are in line for an intervention and in what order ;
- a color frame which identifies the « speaker » participant ;

Perspectives

We are now in the process of experimenting the *Virtual Moderator*™ system in the context of a distance education program where students have to collaborate in various tasks. We also plan to use the *Virtual Moderator*™ to conduct researches on how students interact and learn using such tools. In both cases, those activities should result in communications being prepared for a forum like this one.

Interactive Computer Ethics Explorer

Dr. Walter Maner
Bowling Green State University
Department of Computer Science
Bowling Green, Ohio

For almost all of recorded history ethical issues have been decided according to neighborhood, community, or national norms. Today the Internet breaks these geographic barriers and forces us, for the first time, to deal with information ethics on a global scale. Community standards still prevail; the difference is that the "community" has become the world.

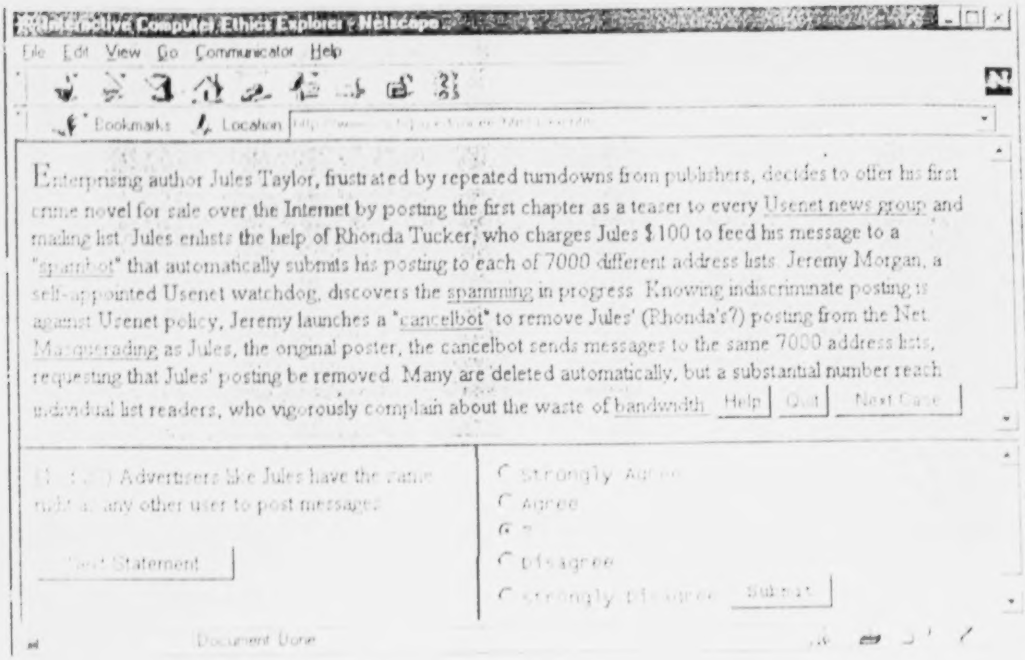
Given this new reality, netizens may wish to explore ethical issues within a framework that will allow them to learn immediately how their own opinions compare to those of people from all over the world. This is the concept behind the Interactive Computer Ethics Explorer (ICEE).

The working prototype invites users to explore the ethical issues surrounding a selected case (e.g., Internet spamming), reveal their own opinion in response to one of twenty focusing questions, and then immediately discover the positions other world citizens have taken on the same question. Demographic data are presented in the form of a bar graph generated on-the-fly. Comparison data initially includes everyone but can be restricted by the user to include only males, females, people under 30, people over 30, US residents, or residents of other countries. Only the 100 most recent responses are saved to use as the basis for further comparisons, so visiting ICEE a second time will likely produce different results and give users even more to think about.

In a traditional scientific study, survey administrators would take steps to prevent survey takers from knowing how other participants have responded until long after the survey is complete. ICEE, on the other hand, does not aim primarily to generate statistics but rather aims to create an opportunity for reflective moral self-development. Because moral growth has always had a cooperative and social dimension, it is often important to explore ethical issues in real time, in interaction with other thoughtful persons. With ICEE, we can extend this social dimension to any corner of the world touched by the Internet.

ICEE establishes an interaction paradigm that has application well beyond ethics. Instead of an ethical case study, one might substitute the text of proposed legislation, a work of art, or a design proposal. Its domain includes any idea that becomes more meaningful when those exploring it are confronted by the opinions of other people.

ICEE is currently implemented in HTML, Javascript and Perl, and makes appropriate use of frames. Typically, the screen is divided into three panels.


Interactive Computer Ethics Explorer - Netscape
File Edit View Go Communicator Help
Bookmarks
Location
Help
Quit
Next Case
Next Statement
Submit
Document Done

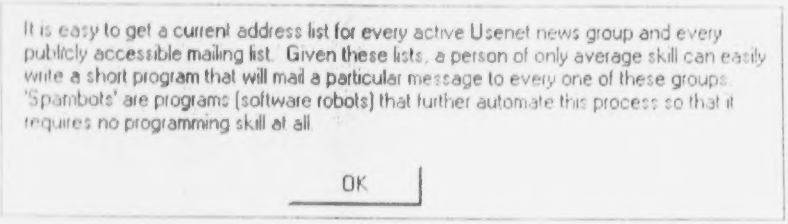
Enterprising author Jules Taylor, frustrated by repeated turndowns from publishers, decides to offer his first crime novel for sale over the Internet by posting the first chapter as a teaser to every Usenet news group and mailing list. Jules enlists the help of Rhonda Tucker, who charges Jules $100 to feed his message to a "spambot" that automatically submits his posting to each of 7000 different address lists. Jeremy Morgan, a self-appointed Usenet watchdog, discovers the spamming in progress. Knowing indiscriminate posting is against Usenet policy, Jeremy launches a "cancelbot" to remove Jules' (Rhonda's?) posting from the Net. Masquerading as Jules, the original poster, the cancelbot sends messages to the same 7000 address lists, requesting that Jules' posting be removed. Many are deleted automatically, but a substantial number reach individual list readers, who vigorously complain about the waste of bandwidth.

Advertisers like Jules have the same right as any other user to post messages.

○ Strongly Agree
○ Agree
● ?
○ Disagree
○ Strongly Disagree

Figure 1: Typical ICEE Screen

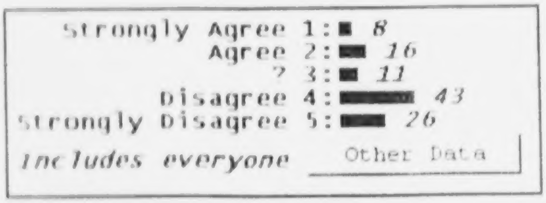
The large upper panel contains a miniature case study (ethical scenario). Various words and phrases contained in the scenario are hot-linked to pop-up windows containing additional explanatory information, like this one for "spambot".



It is easy to get a current address list for every active Usenet news group and every publicly accessible mailing list. Given these lists, a person of only average skill can easily write a short program that will mail a particular message to every one of these groups. 'Spambots' are programs (software robots) that further automate this process so that it requires no programming skill at all.

Figure 2: Popup Explanation of "Spambot"

The second panel, on the lower left, contains a focusing statement based on the scenario, to which the user responds by clicking one of the radio buttons in the third panel displayed on the lower right. Next users click the [Submit] button, whereupon they have a chance to see how the last 100 people responded to this particular statement, themselves included.

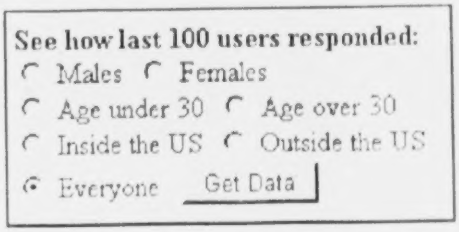


```
   Strongly Agree 1:■ 8
            Agree 2:■■ 16
                ? 3:■ 11
         Disagree 4:■■■■■ 43
Strongly Disagree 5:■■■ 26
Includes everyone
```

Figure 3: Realtime ICEE Statistics

BEST COPY AVAILABLE

Requested data are drawn in real time as a bar graph. Various demographic breakdowns are offered as options.



See how last 100 users responded:

○ Males ○ Females

○ Age under 30 ○ Age over 30

○ Inside the US ○ Outside the US

● Everyone

Figure 4: ICEE Demographic Choices

Finally, the user moves on to the next statement in the series by clicking [Next Statement] or selects another case study to explore.

The current version of ICEE has undergone extensive usability testing. An improved version will debut at WebNet97. Most notably, the new version makes it easy for non-technical people to create new content for display within the ICEE framework.

BEST COPY AVAILABLE

A Group Interaction System for Web-Based Market Places

Peter Manhart
Karin Schmidt
Johannes Bumiller
Felix Meyer
Computer Supported Teamwork
Daimler-Benz Research and Technology
Germany
{manhart, karin.schmidt, bumiller, felix.meyer}@dbag.ulm.DaimlerBenz.COM

Introduction

Today teleshopping can be performed on various electronic markets beared by different technologies e.g. through television, videotext, via online-services and more recently based on the WWW. All these scenarios are characterized by a very low or even no interaction. Mechanisms for interaction are mainly based on email, fax, or telephone ([BEV96], [NSNS97]). However the mechanisms of a real market are highly interactive and determined by offer, demand, and negotiation.

Therefore concepts are needed to map or even improve the communication and interaction of traditional markets. Beside the „pure shopping from a list" three main aspects are important to consider: (1) Awareness of other market participants, (2) means for interaction, e.g. via communication, and (3) building of groups e.g. for reasons of enhanced communication, or joint shopping, or for cooperative navigation through the market system. There exist propositions for interaction in virtual rooms ([RG96], [FKM96]), but all these tools are not specifically tailored for electronic markets.

This paper shortly describes the development of a WWW-based GroupInterAction (GIA) system which provides a more interesting shopping experience through a social environment. In its basic form the market place with its participants is enriched by dialogs of individuals with a supplier or another specialized partner. The next level in the GIA system realizes a perception of other market participants and their activities using VRML. The highest level supports different possiblities for group building e.g. for location chat, acquaintance chat, or even cooperative navigation. The functionality of cooperation and coordination is needed.

The GIA System

Whereas a „shopping from the list" scenario can be implemented on the basis of human computer interaction a lot of other shopping scenarios are characterized by the need for interaction between humans. „Window Shopping", „Group Shopping", „Non Expert Shopping", and „Unexperienced Web User" are examples of social shopping scenarios that are not supported by existing electronic commerce systems and motivated us to work on a innovative component of future electronic malls: A GroupInterAction system called GIA.

Model of Human Interaction and GIA Basic Functionalities

A scenario-based analysis of human interaction on markets resulted in a layered model of human interaction. The lower three layers are implemented in existing markets: (layer 1) pure representation of the market objects (layer 2) human computer interaction (HCI) with the market objects, and (layer 3) dialog communication in separated „chat rooms". In order to support extensive human interaction on electronic markets it was necessary to enhance the existing functionalities by (layer 4) anonymous interaction between market participants in parallel to HCI and (layer 5) human interaction within closed social groups.

In order to support the CCSW-oriented layers (layers 3-5), four basic functionalities of GIA were identified:

(1) „User Perception" enables the market participants to perceive each other. The visualization of the other users is an important prerequisite for group interaction. (2)„Location Chat" combines dynamic group building based on user´s actual location (e.g. a logical cluster of web pages) with chat functionality within the location group. (3)„Acquaintance Chat" combines building of closed social groups (acquaintance groups) with chat

functionality. (4)„Cooperative Navigation" allows the users to link themselves to other users and to jump to an other user´s location without knowing or typing the URL of his location. The implementation of these basic functionalities fullfill much of the goals derived from the layered model of human interaction and turns the electronic market from a single user application to a social place.

The GIA System Architecture

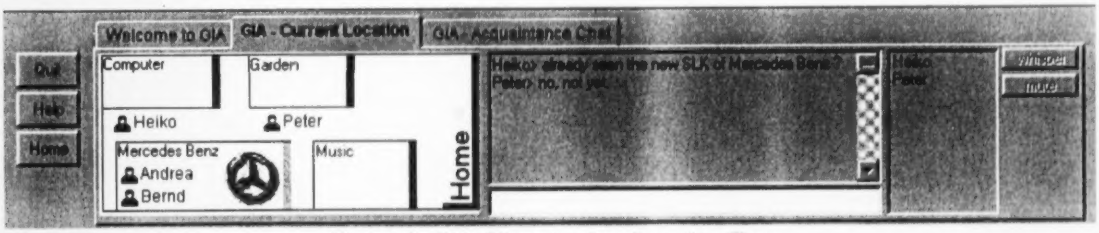
The GIA prototype is based on a client server architecture, that is characterized by the following design goals: Efficient and easy upgrade of existing web based electronic malls with group interaction functionalities, scalability, and stability.

The GIA server runs as an additional service on an internet server and is structured into three cooperating managers: the client manager (client information), the group manager (dynamic grouping) and the location manager (visualization maps for user perception). Only few changes to the web server, that furthermore manages the pure market representation, are required.

The GIA client is implemented as an JAVA applet and mainly consists of data management and visualization widgets, namely a room map panel, a chat panel and a panel for cooperative navigation (see figure 1). The market participant still uses a standard web browser for accessing the market.

The GIA prototype

A GIA prototype has been developed to demonstrate our group interaction concepts, to evaluate the usability, and performance. When loading the entry page of the demonstrator the browser splits into two horizontal frames. In the upper frame the GIA-Client runs as a Java applet (see screenshot in figure 1). In the lower frame a page of the market is loaded. On the left part of the GIA client applet, the map provides a room overview to the user's current location. The „Room Map Panel" visualizes not only passive market objects, but also other visitors. On the right part of the GIA client applet, the „Location Chat Panel" is located. It enables the user to communicate with other people in his current environment. As soon as the user moves to a different location, his environment changes (Location Chat Group and Room Map).



Heiko> already seen the new SLK of Mercedes Benz ?
Peter> no, not yet.

Figure 1: GIA Client Applet - Location Chat

Conclusions and Ongoing Work

Within a standard internet environment the presented GroupInterAction (GIA) system augments electronic markets with advanced and social functionalities (e.g. user perception, location- or acquaintance chat, and cooperative navigation). The prototype runs stable in a real-life environment and is able to handle group sizes of up to 20 persons each satisfyingly.

The ongoing work focusses on one hand on enhancing the GIA system with additional functionality for more structured forms of group interaction and more generic group building mechanisms and on the other hand on the use of the GIA system in an electronic market.

References

[BEV96] *Blackburg Electronic Village*, Virginia Tech, Bell Atlantic of Virginia; http://www.bev.net.

[NSNS97] New South Network Services. *The Hall of Malls*. http://nsns.com/MouseTracks/HallofMalls.html

[RG96] Roseman, M., Greenberg, S. (1996). *TeamRooms:Network Places for Collaboration*. ACM Conference on CSCW 96, Boston, MA.

[FKM96] Fitzgerald, G., Kaplan, S., Mansfield, T. (1996). *Physical Spaces, Virtual Places and Social Worlds: A study of work in the virtual*. ACM Conference on CSCW 96, Boston, MA.

BEST COPY AVAILABLE

Searching for Intelligence on the Web: Exploring the use of Multiple Intelligence Theory in the Internet Environment

Ana Beatriz Martinez, Ed.S., M.Ed.
University of Arkansas, Fayetteville
Adult & Vocational Education, Grad.Ed. Bldg. 117
Fayetteville, AR., USA 72701
501-575-5100
martinez@comp.uark.edu

Kristina Ann Fritts, M.Ed.
University of Arkansas, Fayetteville
Curriculum & Instruction, Grad.Ed. Bldg. 339
Fayetteville, AR., USA 72701
501-575-5100
kfritts@comp.uark.edu

Introduction

Learning is no longer restricted to structured classrooms, television, or even distance education. The classroom has become global with students numbering in the millions, and speaking many languages. These learners also access, utilize, and manipulate information in different ways. The 'information superhighway' is now assuming the role of trainer, tutor, and facilitator of learning. This instrument can take on any number of configurations with regard to design, style, and format. Sensitivity to the Multiple Intelligences Theory, as developed by Howard Gardner, can help designers and instructors to formulate different strategies for content delivery. Optimizing on these intelligences can be one of the keys in the communicative process.

Purpose

The purpose of this paper is to discuss the use of Multiple Intelligences, and provide an example of how to effectively utilize this theory in web-page development. We would like to challenge web visitors and instructional designers to offer several routes to mastery of information.

Pedagogical Factors

The access to information plus the possibilities of interaction in real time, allow designers to explore different educational possibilities. In fact, the Internet permits the use of a vast number of tools to create new spaces for learning, yet different from the classical classroom. Virtual classroom, teleconference, and the developing of cooperative learning are only some of the multiples possibilities that have been explored. In order to make the Internet a more intelligent technology the following elements must be considered when designing Internet pages:

* Content
* Format
* Users

The Seven Intelligences

Gardner (1992) identifies at least seven types of intelligence. Linguistic or Verbal Intelligence is related to words and language, written and spoken. Logical-Mathematical Intelligence deals with reasoning and the recognition of abstract patterns. Spatial Intelligence refers to the sense of sight and the ability to create internal images or pictures, but also the sense of spatial orientation. Musical Intelligence is based on the recognition of sounds and rhythms. Bodily-Kinesthetic Intelligence is related to physical movement and motion, but also with awareness of the body and it's needs. Interpersonal Intelligence is based on person-to-person relationships and communication. Finally, Intrapersonal Intelligence deals with self concept, metacognition, and spiritual realities.

Each individual possess each of these skills to some extent, but they differ in the degree and the nature of their combination. On the other hand, each one is universal and independent of educational or cultural differences. Multiple Intelligences are more generic than learning styles and they are not necessarily related to the content. Sensitivity to Multiple Intelligences may help designers to determine better ways of presenting content, and optimize the communicative process.

Based on Gardner's Multiple Intelligences Theory we can make the following assumptions:
* Multiple Intelligences are universal.
* Multiple Intelligence are independent of education or cultural influence.
* Everyone has Multiple Intelligences but in different degrees and combinations.

Designing a Web-page based on Multiple Intelligence

The Internet's primary function is to deliver information in a fast and efficient manner. The web is an extraordinary media which allows for use of many different resources in the design of pages. Text, video, audio, are not isolated resources if we can think of them related to the learner's or users needs. In that sense it is possible to consider different resources to reach particular intelligences or the same resources but with different functions. Incorporating the seven intelligences can ensure the attention of all users, regardless of how they learn. For example, musical learners can be stimulated by sounds or the human voice, whereas an interpersonal learner can be reached by interactive stimulous.

Sensitivity to Multiple Intelligences may help designers and instructors to formulate different strategies for content delivery, optimizing these intelligence which may be key to the communicative process. The true challenge for web designers on any level is to offer several routes to mastery, and to increase the likelihood that any individual can attain knowledge in the process of learning.

References

Gardner, H. (1983). Frames of Mind. New York: Basic Books.

Gardner, H. (1992). Multiple Intelligences: The Theory in Practice. NY: Basic Books.

Hotte, R. (1993). Encadrement assiste'e par ordinateur et formation a'distance. Revue de l'e'ducation a'distance, VIII (2), 37-53.

Lazear, D. (1991). Seven Ways of Knowing. Palatine, Illinois: Skylight.

Levi, P. (1994). L'intelligence collective: Pour une anthropologie ducyberspace. Collecton Sciences et societ'e. Editions la d'ecouverte, Paris. 234.

Enhancing Commercial Web Sites by Dynamic User-Interaction-Management

Felix Meyer
Computer Supported Teamwork
Daimler-Benz Research and Technology
Germany
felix.meyer@dbag.ulm.DaimlerBenz.com

Introduction

Electronic Commerce changes communication between companies and their customers. On a customer friendly Web site a visitor should not be forced to find his way to the desired information all by himself. The company represented by its Web site should ask the customer what it could do for him. For information that exceeds "customer self-service" it is necessary to tie the customer into internal processes of the company. Globally collecting incoming feedback and answering it in general terms puts a company in a bad perspective. It is the quality of these services which adds the competitive extra value to a company's Web site.

This paper describes a system to help companies to react to user feedback in a structured and competent way. The "User-Interaction-Management System" maps incoming feedback to company experts and thus ensures competent answers and satisfied customers. It watches over reaction time and reminds experts to answer if the answer is over due. Through every answer the system's mapping mechanism is trained to better classify experts and feedback. It is developed for use on any Web-system with least integration effort.

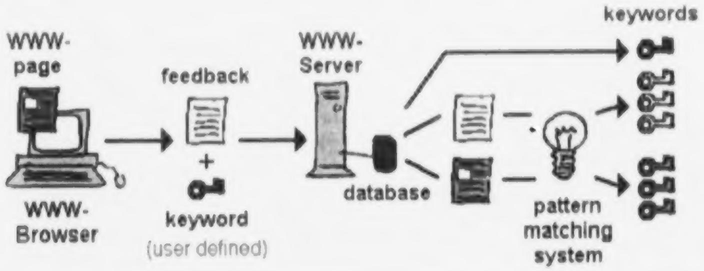
Concept of User-Interaction-Management

The User-Interaction-Management System ties dynamic interaction of customers on a company's Web site into the company's internal work processes. Intentionally given or unintentionally collected feedback is directed through a workflow to the responsible people within the company.

User interaction is defined as a request for information involving people on the company's side. This includes product information, on-line consulting and service. A simple database query is not within this scope.

Determination of Keywords

Incoming feedback has to find its way to the best expert for a competent answer. There are three sources of information that help the system to find the expert. The first source is the feedback itself, the second source is the specific WWW-page where the feedback originated from and the third source is an optional keyword that the customer can add to his feedback. As soon as feedback comes in the workflow system is triggered and hands feedback and WWW-page over to a pattern matching system to find keywords. The user defined keyword is carried along the process without change (Figure 1).



BEST COPY AVAILABLE

Figure 1: Determination of Keywords

Keyword Judging by Expert

Matching between expert roles within the workflow system and keywords is done by a weighted "knowledge metric" - a mechanism which computes the relationship between the keywords given by the pattern matching system and the expert knowledge - represented through weighted keywords - within the company.

To improve this matching mechanism the knowledge metric is constantly trained. Every expert answer changes the weight of a keyword-expert-relationship. After answering an expert has to judge how well the feedback fit his knowledge and how well the keywords from the pattern matching system matched the actual content of the feedback and the WWW-page. The expert can even cross out keywords if they didn't describe the content well. After each answer both the WWW-page and the feedback are stored in a database and closely associated with their keywords (Figure 2).

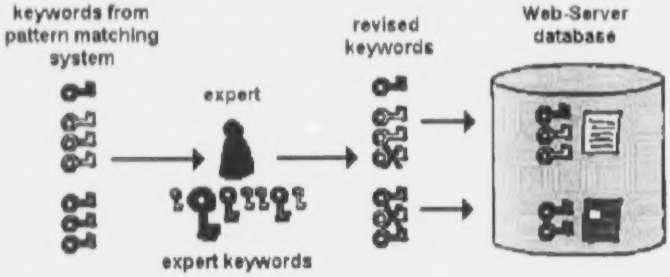


Figure 2: Keyword Judging by Expert

Dynamic Workflow Generation

From the keywords the User-Interaction-Management System dynamically generates a workflow which guarantees a reply to the customer within a certain time frame. It can also direct feedback in parallel or one after another to different experts for in depth verification or multiple aspects of the problem. Experts can reject feedback to the workflow system if the matching wasn't accurate. In this case the workflow system would search for the next best expert and forward the feedback to him.

To support the expert the User-Interaction-Management System also passes on information about the customer (if known from previous interaction) and the WWW-page where the feedback originated. Through keyword search links to previous feedback and answers with similar topics are offered by the system. This gives the expert an advanced support for understanding the problem and giving helpful information to the customer.

Implementation of Prototype System

The implementation of the User-Interaction-Management System is platform independent and can extend the functionality of any running Web site. Only standard interfaces to the Web server are used. Any Web-Browser supporting Java Script is able to run the client side scripts.

Just like the customers the internal experts also access the User-Interaction-Management System through a Web-Browser. The underlying Oracle database is accessed through the Web Request Broker which then executes stored procedures on the database. Since the workflow is not very complex the workflow engine is a custom implementation which is based on the same database.

Conclusion

BEST COPY AVAILABLE

Looking around in the Internet shows a great demand for a tool to handle customer feedback in a structured and reliable way. A dynamic workflow system to carry external feedback into internal processes is the first approach. As work on the User-Interaction-Management System is still ongoing a prove of concept through a field study is still missing. A business unit within the Daimler-Benz group is already interested to provide a test bed.

Customizing the Content of Courseware

Uwe Neuhaus, Gunter Schlageter
Department of Computer Science
Distance University of Hagen, Germany
E-mail: Uwe.Neuhaus@FernUni-Hagen.de

Our research group at the Distance University of Hagen is building a "Virtual University", i.e. a network-based system for distance education that encompasses all aspects of a university (see [Buhrmann et al. 1996].) Within this project the individualized structuring of the information-space plays an important role. Ideally, all the information within the Virtual University is presented to a student in such a way that it serves his or her individual needs best. One aspect of individualization is the customization of courseware. Curently, we are developing a system of modular courses which eventually will utilize the large stock of courseware already contained within the Virtual University. This paper gives a brief overview of this aspect of our present research:

To customize the content of courseware means to compile for each learner an individual course that takes into consideration the learner's specific interests and previous knowledge of the subject. Therefore, the underlying idea is to subdivide courses in (relatively) independent modules that can be combined according to the specific situation of the learner. The learner must be able to include all course-modules that he or she is interested in or needs to know about, and exclude all others. Furthermore, it must be possible to add additional modules and exchange existing ones at a later point in time. At any stage the selected modules should form a homogeneous whole. The problems that the development of such modular courseware poses (as well as some possible solutions to these problems) are described below:

Inter-Module Dependencies

While it is not possible to subdivide all forms of texts in (more or less) independent units, many textbooks can be (and often are) written in such a way. How small such independent units can be depends to a high degree on the subject matter. The division into chapters and subchapters usally can serve as a guideline. Often sections of a course will depend (i.e. presuppose the knowledge of) preceding sections. Dependencies that go beyond the borders of a course module have to be made explicit. In this way every module can be assigned a (possibly empty) set of preconditions that have to be fulfilled (either by working through the required modules or by stating that one has already comparable knowledge) before the module is studied.

The Selection Process

How do learners decide which modules serve their needs best? For learners who have already a basic knowledge of the subject area it might suffice to offer short descriptions of the content and the preconditions of the modules. However, for inexperienced learners more sophisticated mechanisms are needed. Simply to offer standard selections for certain topics is an unsatisfactory solution because in this way some of the advantages of modular courses are lost. A better way is to start with a module that gives an overview of a certain subject area (this module should have an empty set of preconditions) and then, bit by bit, add more modules corresponding to the learner's developing expertise and interest.

Integration of Modules

After the desired modules are selected they have to be made into an integral whole. If the modules are not smoothly integrated, they will merely be a collection of disconnected pieces of information. The integration of modules affects the logical structure and the layout of the documents contained in the modules as well as the navigation within and between these documents.

- *Logical structure:* An individually compiled course needs to have a common table of contents, a common index, a common glossary and a common bibliography. Therefore it is necessary to have access to the logical structure of the documents in the individual modules. This can be achieved if the authors of the modules use a common markup language. The Extensible Markup Language (XML) – a recently developed open standard that is particulary suited to specify markup standards for web documents (see [W3-Consortium 1996] and [Bosak 1997]) – offers the needed functionality and flexibility to define document types of the required kind. It can serve as a basis on which the functionality needed for common tables and indices can be built on.
- *Layout*: All modules within an individually compiled course should have the same layout ("the same look and feel".) This can be achieved if the information about the layout of the documents contained in the different modules is kept seperately from the content and the logical structure of the documents. Under these circumstances the layout of a document can be changed afterwards, and therefore the modules can be given a uniform appearance. A technical solution to achieve a homogeneous layout across all modules can be based on the usage of a style sheet language like the Document Style Semantics and Specification Language (DSSSL), a standardized stylesheet language that is currently under development (see [DSSSL 1996]).
- *Navigation:* Navigating from page to page should be consistent throughout the course. Therefore, a navigation interface has to be designed into which a common navigation unit can plug in later. A related problem is that of "semi-external hypertext links", i.e. hyperlinks that point from one module to another. When modules are compiled into a course, when they are later on replaced or supplemented by further modules, the existing external hyperlinks have to be taken care of. The functionality offered by HTML – the current standard for web-documents – is by far not sufficient for the described purposes. A much more powerful hypertext mechanism is needed, allowing for – among other things – bidirectional and typed links.

Possible Fields of Application

Modular courseware can be useful not only for distance education but also for regular universities. Lecturers, for example, can select modules for an introductory course, an in-depth course or a specialized course. Given the high development cost of multimedia courseware, the reuse of existing modules can help dramatically to reduce the expenditures for learning material. Lecturers utilize existing material but are still free to choose what best serves their purposes. Furthermore, the concept of modular courses is also very useful in the rapidly growing field of continuing education, because here the learners' previous knowledge of the subject as well as their needs and interests can vary widely.

References

[Bosak 1997] Bosak, J. (1997). *XML, Java, and the Future of the Web*. [`http://sunsite.unc.edu/pub/sun-info/standards/xml/why/xmlapps.htm`].

[Buhrmann et al., 1996] Buhrmann, P., Mittrach, S., & Schlageter, G. (1996). The FernUniversität as a Virtual University – Concepts, Experiences, Developments. *Online Educa*, Berlin, Nov. 1996.

[DSSSL 1996] Joint Technical Commitee ISO/IEC JTC 1. (1996). *Document Style Semantics and Specification Language* (Final Draft). [`http://occam.provo.novell.com/dsssl/dsssl96/1.toc`] .

[W3-Consortium 1996] W3-Consortium. (1996). *Extensible Markup Language (XML)*. [`http://www.w3.org/XML/`] .

Acknowledgements

The results presented in this paper have been developed within the framework of the project „Virtual University" at the Distance University of Hagen. It is funded by the federal state Northrhine Westfalia and supported by Oracle, Sun Microsystems and Sybase. Project coordinators are Prof. Dr. G. Schlageter and Prof. Dr.-Ing. F. Kaderali.

Access to Web-based Special Education

Kevin K. Nguyen
Adaptive Technology Resource Centre, University of Toronto
Toronto, Ontario, Canada
k.nguyen@utoronto.ca

Although, web-based distance education programs address geographical and cost barriers, they usually ignore access barriers to students with special needs (i.e. those with sensory, motor or cognitive disabilities). Distance education programs should ensure that conduits, and not barriers, to information are created. When planning a web-based special education program the following concerns should be considered: how to increase Web access to persons with disabilities by addressing access issues on both the client and the service side; how to optimize the use of innovative web technologies to transmit interesting yet accessible learning materials; how to increase community amongst special education students and teachers.

Web Access and Special Needs

A web-based special education infrastructure holds promise for opening up new windows of opportunity for students with special needs. For example, because web page content usually consists of electronic text, a blind student can use a screen reader to audibly present and navigate the information. The font of most web browsers can also be easily increased for persons with low vision. However, aside from some of the inherently accessible properties of the Web, many barriers can be unnecessarily created on both the client and the 'service' side. The phrase *client-service*, instead of *client-server*, is emphasized because providing education is a service, and all services should accommodate persons with disabilities.

On the client side, many types of adaptive technology exist, including alternative mouse systems, alternative keyboards, voice recognition systems, refreshable Braille displays, and screen readers. These systems make it easier for persons with disabilities to access their computer and the Internet [Nguyen and Petty 1997]. There are several World Wide Web browsers available that vary in their accessibility features. Browsers can have keyboard equivalents for hypertext links, frame navigation and built-in alternative display modes. When creating a distance education infrastructure, it is best to create a site that is browser-independent and to avoid the use of proprietary browser features or custom HTML tags. Forcing everyone to adhere to a single type of browser is not optimal for the diversity within the special needs population.

Just as wheelchairs can only function if a flat surface is available, client-side access systems can only work if small provisions on the service side are present and barriers are not erected. Simple, transparent web access provisions include alt-text, text equivalents to image-links, and standardized navigation schemes [Letourneau 1996]. Inaccessible web design can block access to information for someone using adaptive technology [Treviranus and Serflek 1996]. For example information embedded in bitmapped text or images without appropriate alt-text will be missed by people using Braille displays, screen readers and text browsers. Hyperlinks embedded within the bullets of a list (instead of in the list items) may be too small for someone using an alternative pointing device to target, and too indescript for a person using a screen reader to differentiate. Overuse of frames and tables can unnecessarily increase the complexity of the page for persons with and without learning disabilities.

Enriching Access through Innovative Web Technology

Innovative web technologies can now be used to deliver inexpensive, on-demand and interactive teaching materials. Streaming video and audio can be used to increase the richness of information transferred over the Web. Captioned video can be created for those who are deaf. Descriptive audio can be used to provide an unobtrusive narration of video for persons who are blind, and text transcripts of video or audio clips can help anyone using a Braille display. The multi-modal availability of resources can supplant some students needs for adaptive technology. For example, a blind student can simply listen to a live audio stream of a news report instead of having a screen reader speak a transcript of the broadcast.

Real time web based videoconferencing, audio conferencing, collaborative work areas, chat rooms or IRC, messaging boards and email can all be used for creating interactive learning environments. New web technologies are constantly being introduced into the market. However, each technology should be justified in terms of cost, target audience and accessibility before implementing them within a distance education program. Sometimes "less is best". For example, a teacher discussion forum about an article on ADHD can occur either through web-based videoconferencing or email. The former requires tight scheduling as well as a video capture card, a camera, and adequate Internet bandwidth for each participant. Issues involving moderation, time zone discrepancies and access can arise with video conferencing. E-mail discussions however have no special hardware requirements, have few or no access problems and can occur asynchronously. Participants can compose their thoughts and replies with as much time as required. There is evidence that email may even allow deeper and more open communication because of a veil of anonymity [Gold 1997]. Each method has its advantages and disadvantages, but both allow interactivity among peers.

Generally, there is a need for enticing, current curriculum for special needs students and teachers. It is essential that curriculum is motivational so that student will *want* to learn. Proprietary conversions of resources from education-related centres such as, zoos, science centres, museums and art galleries, can be created specifically for special needs students and placed online. Small collaborative projects between such organizations, teachers and distance education administrators can target resources that are lacking in the field. Once a teaching module is created in digital form, it can easily be updated, modified, reused and distributed amongst the special education community.

Community Building for Student and Teacher Support

A successful technology infrastructure cannot be implemented by simply installing computers, software and network connections into the classroom. It is important that teachers know the capacity and limit of the technology. But do teachers have fears that students will know more about the technology than they will? Some teachers may initially oppose the change and feel that they must struggle to stay on top of the technology in order to teach it to their students. However this should not be the case. The technologies are simply tools for augmenting information access and communication. If a quadriplegic student uses a voice recognition system in order to compose his or her writings, the teacher should see the technology as being no more intimidating than paper and pen. If a student is attending a class via videoconferencing, the teacher should not see a "videoconferencing camera and computer system", but view it as a student attending class. The technology should never overshadow the individual, and should eventually become transparent.

Teachers, as well as students, require access to special education resources [Baker and Danley 1996]. The advantage of web-based special education is that the mechanism which allows students to access curriculum also allows teachers to access support resources and to communicate with other colleagues. Providing classrooms with a web-based program can broaden the type and scope of information both special education students and teachers can access. Global resources, expertise and examples of best practices on special education can be easily shared [Paulet 1989].

One of the goals of any web-based distance education infrastructure is to increase learning communities and to expand the regular confines of traditional classrooms. By linking together several special education classrooms with web-based telecommunications and multi-user environments, students and teachers from different schools will be able to share resources, collaborate towards goals and communicate with others who share common interests and concerns [Gold 1996]. For example an autistic child in a rural community who communicates through BLISS symbolics only with her parents and teachers may now be able to reach out and "talk" to other children using BLISS via a web medium.

Personal tutors are sometimes used for children who must stay at home due to medical reasons. One disadvantage to this strategy is that the tutor cannot provide group interaction between fellow peers, something that is important in the social-educational development of a child [Williams et al. 1995].

Interactive windows into the classroom can be opened for isolated students through web-based media and videoconferencing.

Special Needs Opportunity Window (SNOW)

The SNOW Project is a one-year pilot project to create a province-wide distance education web-infrastructure for special education teachers and students aimed at enhancing literacy and numeracy in the primary grades. This work-in-progress hopes to provide an ideal model of web-based distance education that meets the unique requirements of both special education teachers and students. An information resource, community network and videoconferencing system will all be built upon a web infrastructure. Access technology will be placed in several special education classrooms throughout Ontario. Special education courses and workshops for pre-service and in-service teachers will also be available through SNOW. SNOW incorporates and addresses many of the issues discussed in this paper. It is hoped that by bringing these issues to the public, other web-based distance education programs can make accommodations for those with special needs and in doing so adhere to universal design principles. To find out more about SNOW visit http://snow.utoronto.ca or contact the author.

Acknowledgements

This project was made possible through funding from the Ontario Ministry of Education and Training's Technology Incentive Partnership Program (TIPP96-082). Acknowledgment goes out to all on the SNOW team including Greg Gay, Laurie Harrison, Jim Lam, John Lubert, and Tim Higgins. Special thanks to Jutta Treviranus.

References

[Armstrong 1995] Armstrong, K. (1995). Special Software for Special Kids. Technology and Learning, 16(2), 56-61.

[Baker and Danley 1996] Baker, C. and Danley, B. (1996). Comparing Computer-Assisted Instruction and Traditional Instruction for Preparing Regular Teachers to Serve Students with Disabilities. Computers in the Schools, 12(4), 31-38.

[Gold 1996] Gold, J. (1996). Enhancing Professional Development in Special Education Through the Web. Computer-Mediated Communication Magazine, Oct 1996. (http://www.december.com/cmc/mag/1996/oct/gold.html)

[Gold 1997] Gold, J. (1997). Does CMC Present Individuals with Disabilities Opportunities or Barriers? Computer-Mediated Communication Magazine, Jan 1997. (http://www.december.com/cmc/mag/1997/jan/gold.html)

[Letourneau 1996] Letourneau, C. (1996) Accessible Web Page Design. Starling Access Services. (http://www.igs.net/~starling/acc/index.htm)

[Nguyen and Petty 1997] Nguyen, K. and Petty, L. (1996). Web Browsing through Adaptive Technology: An Information Resource for Consumers. The Sixth International World Wide Web Conference, 1997, International World Wide Web Conference Committee, Santa Clara, CA. CD-ROM Proceedings.

[Paulet 1989] Paulet, R. (1989). Assessing a Rural Special Education Programme. Educational Research, 31(2), 125-33.

[Treviranus and Serflek 1996] Treviranus, J., Serflek, C. (1996). Alternative Access to the World Wide Web. Adaptive Technology Resource Centre, University of Toronto. (http://www.utoronto.ca/atrc/rd/library/papers/WWW.html)

[Williams, Fels, Treviranus, Smith, Spargo and Eagleson 1995] Williams, L., Fels, D., Treviranus, J., Smith, G., Spargo, D. and Eagleson, R. (1996). Control of a remote communication system by children. CybErg 1996, Curtin University, Perth Australia.
(http://www.curtin.edu.au/conference/cyberg/centre/paper/williams/paper.html)

Connecting Business and Education Through the World Wide Web

Gian Pagnucci
Department of English
Indiana University of Pennsylvania
United States of America
pagnucci@grove.iup.edu

William Macauley
Department of English
Indiana University of Pennsylvania
United States of America
tgzd@grove.iup.edu

There has long been a conflict between business and education. This conflict isn't always clearly defined, often existing simply as mutual suspicion. Business executives may fear that teachers are not trying to prepare students for work, and teachers worry that preparing students only for the workplace limits the depth of students' educations. This conflict is one which we have found can be resolved.

We teach introductory college writing. We want our students to learn to be creative, critical, and aesthetically aware. But we also are deeply concerned that our students learn how to write on the job, meet deadlines, set schedules, negotiate challenges, work on teams, and learn how to manage responsibility. For us, teaching writing is not telling students one more time how to write an essay about summer vacation.

Like most college writing classes, our course requires students to write essays, stories, poems, opinion pieces, and critiques, all traditional ways of helping students become skilled academic writers. To connect our college writing course more directly to professional concerns, though, we have also adopted four strategies which link students' academic work and future professions. We have students learn common workplace technologies, study and practice traditional workplace documents, develop management skills, and investigate their future careers. By employing these strategies, we believe we enable students to be effective workers as well as good students and thereby ease the disparity between the academy and the workplace.

Technology can be a blessing or a curse. When teachers rush to add new technologies to their courses simply for the sake of using what's new, these technologies often end up being more burdens than useful learning tools. However, when technologies are integrated into a course context in which they are a necessary part of the work, they can enable students to do types of work which were previously not possible. In our college writing course, we've included four specific technologies which writers regularly use to do their professional work: the Internet, e-mail, word processing, and desktop publishing.

Our students use the World Wide Web to do research and use to collect images for doing desktop publishing. They learn how to access the many resources available on the Internet and how to distinguish between those which are valuable to their work and those which lack credibility. The students also learn how to produce their own web pages, manipulate HTML codes, and do writing which will reach beyond the classroom to impact real readers. Along with all this, our students use electronic mail to discuss their web pages and other writing with us and with each other.

Word processing is an absolute essential since hand written and even typed documents are simply unacceptable in the work place and the academy. We expand students' writing skills by enabling them to make use of more advanced word processing features and desktop publishing as well. We help students develop documents with elaborate designs which integrate high quality writing, graphics, and layout features such as headers, borders, and columns. Advanced word processing skills give these students an edge both in school and after. In the work place, being able to make a publication, a newsletter, a web page, or a promotional package will make a student more competitive and marketable.

Along with developing students' technology skills, familiarity with workplace writing is critical. We have students write project proposals and then letters of application to work on these projects. Once students receive the letters of application, they have to respond with letters hiring or rejecting the job candidates. Later, as the students undertake their projects in small teams, we ask them to keep us informed about their progress through email memos, short reports, and thorough project evaluations. This approach helps students develop the ability to work independently, but also gives us as teachers a way to monitor their learning and progress, much as would happen in any workplace. Too often, students write isolated essays or research papers for college writing classes, but our variety of assignments lets them relate one kind of work to another, creating a context for the projects that simulates a professional environment.

This course is a lot of work for us and our students. The amount of time students spend in the classroom is just a fraction of the actual work time they have to devote to course related activities. To handle these intensive work demands, students have to develop their time and project management abilities. We ask students to keep weekly time sheets documenting their class work activities. We have them develop detailed project schedules outlining deadlines, tasks, plans, and progress toward project goals. They also have to figure out how to work with other people, how to use their time wisely, how to negotiate conflicts, and how to apply successfully and efficiently the technologies to which they've been introduced.

Finally, we ask students to research both their working pasts and their future professions. We have students interview workers in their chosen fields, discuss work experiences with their friends and families, and read and write about work environments and trends. This kind of study helps the students gain an appreciation for the kind of work they will be expected to do in the future as well as the skills and abilities they will need to succeed in that work. This process also demystifies the workplace for the student and encourages them to connect directly the work they are doing in school with the work they would like to do after graduation.

By modifying our college writing course, we enable our students to link that work to their own interests and professional goals. It is our hope that this kind of course can uncover the goals of business and the intentions of education. We want to work toward lessening both sides' suspicions. We would be doing our students a great disservice if we did not teach them about the technologies on which the world depends, did not prepare them for the careers they will have, and did not facilitate the lifelong learning they will need to do.

Architecture for Intranet Collaborative Learning Activities

Joaquim Sousa Pinto
University of Aveiro / INESC, 3810 Aveiro, Portugal
email: jsp@inesca.pt
Phone: +351.34.370500 Fax: +351.34.370545

Joaquim Arnaldo Martins
University of Aveiro / INESC, 3810 Aveiro, Portugal
email: jam@inesca.pt
Phone: +351.34.370500 Fax: +351.34.370545

Conceição Baptista Tavares
University of Aveiro / INESC, 3810 Aveiro, Portugal
email: sao@inesca.pt
Phone: +351.34.370500 Fax: +351.34.370545

This paper describes an architecture for a collaborative system to be used in a learning and group training scenario. This architecture is being developed in the aim of the CVMED project, and can be used in several synchronous collaborative scenarios in the education and training fields. The main characteristics of such an architecture are:

- the system does not require a dedicated centralised server;
- the collaborative model is based on replicated tools. Under this collaborative model the whole stations run simultaneously the same applications.
- the system supports several modes of collaboration allowing the implementation of distinct learning scenarios;
- it is based on several tools. Instead of putting the whole functionality within a single application each one covers a specific functionality. According to their functionality, the tools can be classified in three different groups: basic tools (required to operate the system), generic tools (cover generic functionality and are scenario independent), and specific tools (designed according to the scenario where they will be used). This approach is twofold; on one hand it allows a simple development and testing of each system component and the involvement of several groups of developers; on the other hand it allows the development of new specific tools if we decide to use the system in other learning scenarios than the ones for which the system was conceived.
- Since we designed and implemented our system to be used in learning and training scenarios, our specific tools can edit and play hypermedia material.
- besides the development of the specific tools covering predefined learning scenarios, the system allows the integration of single user applications, such as word processors, spreadsheets, ...;
- the target population for the system is medium labs or schools;
- runs in low cost machines, like PC's and uses a window based operating system;
- is scaleable. The minimum configuration requires only two terminal equipments. The system grows only with the connection of new terminal equipments, without special equipment, for example a MCU;
- since it is implemented over stable network protocols, it can be used in several kinds of networks.

To define a work session organizational structure the "*Virtual lobby*" and "*Virtual room*" [Hiltz88, Hiltz94] metaphors were used. The virtual lobby is the system entry point. Within the lobby a user can choose among several virtual rooms controlled by the virtual lobby. The activities developed inside a room can be performed by a single user or by a group of users and can be public or private. When a group of people performs an activity inside a room we use the "conference" metaphor to describe such activity.

In order to make the implementation of the system tools easy a communications protocol, which is responsible for the data transmission among the several conference members, was developed. None of the tools access directly the network. That access is always done through this communications protocol. By the way of this protocol the system tools communicate with each other through the exchange of Protocol Data Unit's (PDU's). To support the system tools the communications protocol specifies several Application Protocol Interface's (API's).

The communications protocol was developed using the TCP/IP protocol family. This allows us to run the system simultaneously in LAN's and WAN's, and to be independent of the hardware manufacturer.

The first prototype was developed in 1993 and 94 [Pinto94, Pinto95]. The first laboratory tests and the carried out evaluation encouraged us for the development of a new version of the system. This second version is implemented with the IP Multicast and the Windows95/NT operating system.

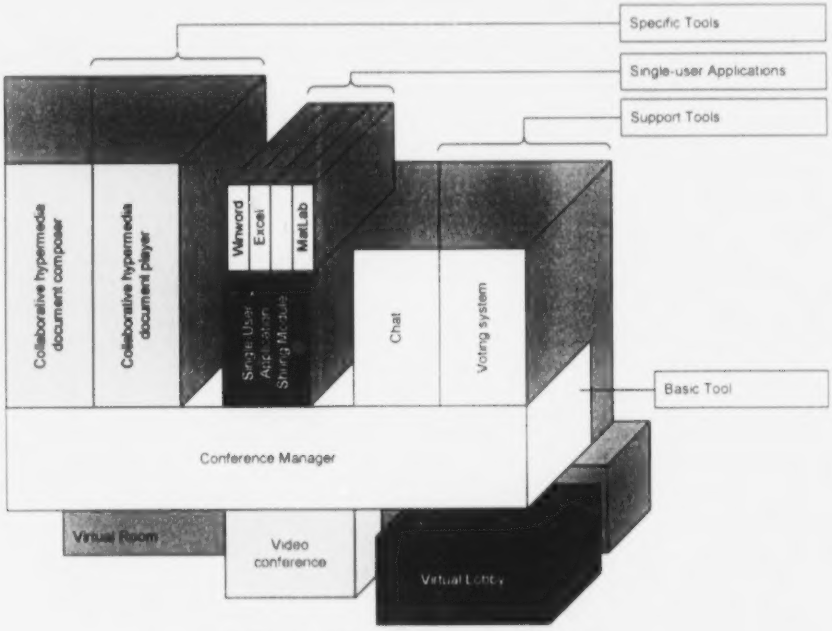

Specific Tools
Single-user Applications
Support Tools
Collaborative hypermedia document composer
Collaborative hypermedia document player
Winword
Excel
MatLab
Single User Application Sharing Module
Chat
Voting system
Basic Tool
Conference Manager
Virtual Room
Video conference
Virtual Lobby

Figure 1 - System General Architecture

References

- *Book references:*

[Hiltz94] Roxanne Starr Hiltz, (1994), "Virtual Classroom: Learning Without Limits Via Computer Networks", Ablex.

- *Proceedings references:*

[Hiltz88] Roxanne Starr Hiltz, (1988), "Collaborative Learning in a Virtual Classroom: Highlights of Findings", *Proceedings of CSCW'88*, ACM, 282-290.

[Pinto94] Joaquim Sousa Pinto, Joaquim Arnaldo Martins, (1994), "'CHEaP': a Cooperative Hypermedia Editor and Player", *Proceedings of IEEE/SBT ITS'94*, Rio de Janeiro, Brazil, August.

[Pinto95] Joaquim Sousa Pinto, Joaquim Arnaldo Martins, (1995), "Design and Implementation of a Virtual Learning Environment", *Proceedings of 1st Cyted-Ritos International Workshop on Groupware (CRIWG'95)*, Lisbon, Portugal, September, 53-64.

Acknowledgements

Part of this work is funded by JNICT (Portuguese R&D agency) project 2/2.1/TIT/1581/95

BEST COPY AVAILABLE

Publishing Complex Network Information on the Intranet

Veikko Punkka
Nokia Telecommunications
Finland
veikko.punkka@ntc.nokia.com

Introduction

Complex telecommunications networks are getting more and more complex every day and various interest groups need various kinds of information of the networks. Network maintenance people need information of the failures in the network and some indicators of the performance of the network, network planners need information of the network usage patterns, customer service people need information of troubles in the network services and coverage, just to mention a few. Not only the content of the information changes, but also its structure and intended audience changes as the networks evolve. Information must be available in a number of locations using a number of access devices with varying capabilities. Only the web can meet all these requirements for information delivery, but the web is not enough. One also needs information collection and publishing systems to make sure the right people get the right information on the right time.
Within the Nokia Network Management Systems the information flow is divided into four phases: data retrieval from the network, data storage, information generation based on the stored data and information distribution. The first three phases that constitute the traditional network management systems are outside the scope of this presentation. These systems are reaching maturity and they satisfy the needs of dedicated network maintenance personnel in the process of keeping the network up and running. Information distribution widens the audience of the network management systems and hence constitutes a significant added value to the network operator.

Information Distribution

The fundamental concept of information distribution is a report. A report is the smallest separately publishable piece of information. Physically it maps to an HTML file. Each report is composed of several smaller sets of information. These may include images made by external systems, forms, links to other reports, Java applets, ... The list of possibilities is endless. The difference between the publishing system and a plain web server is that the publishing system actively knows these parts and builds the reports out of them upon publication. If some of the parts need periodical maintenance, the system knows how to do this and how often to do it. This includes actions such as checking the validity of the links or checking if the data in a spreadsheet component has expired.

Right Information on the Right Time

There are three kinds of reports differentiated by the frequency they are needed. Scheduled reports are generated and published according to a fixed schedule. On demand reports are generated and published when they are requested for the first time. Ad hoc reports are defined on the fly but never published. Scheduled reports are needed regularly or frequently. Examples of such reports are various weekly reports such as reports of certain key performance indicators that are frequently accessed. On demand reports are either based on live information or less frequently needed. Examples of such reports are reports of the current network failures and in depth reports of individual network elements. Both scheduled and on demand reports are based on complete report models that determine what information the report contains and how it is displayed. There are no report models for ad hoc reports, but defining an ad hoc report actually involves first composing a report model and then filling it with information. If the report model seems useful enough the operator can make it either a scheduled or an on demand report.

Right Information for the Right People

Report access is based on user groups. Each user that can access the reports in the system belongs to one or more user groups. Ad hoc reports are generally restricted to expert users that generate report models for other user groups. Scheduled and on demand reports are available to various user groups according to their information content and the needs of the user groups. For each user group the reports are arranged in a hierarchy that supports the conceptual models of that particular user group making it easier for the user to find the information he needs.

As reports are composed of various parts as they are published, the parts may be customized for various user groups. This may include things such as disclosing or summarizing information to a various extent, using terminology familiar to the user group, providing extra reference material for people unfamiliar with certain concepts or even using different languages or character sets for different user groups. This way each user gets the information he needs in a format he can most conveniently utilize.

In addition to providing the users information of the network, the information distribution system provides them report related shared spaces. The users can use these shared spaces to exchange comments and other related information of the published reports. This makes the system also an active collaboration tool. Combined with automated access to electronic mail and short message services it improves the efficiency of the operating personnel. This makes the system cost efficient even for the network operators that do not want to use it to widen the audience of network management systems.

Easy Administration

Easy administration is a vital part of the information distribution system. For this the system must include an administration tool that can perform all the administrative tasks that the system requires. These tasks include definition of reports, management of user groups, report access, navigational hierarchies for the various user groups and connections with external systems. The administrator does not need to be an expert in web administration, but people experienced in the subject area of the reports can carry out the administration without external assistance.

Experience

Within the Nokia Network Management Systems the information distribution system has first been implemented in conjunction of the Network Data Warehouse. It is a system that collects summarized information from various data sources in the managed network and stores it for extensive periods of time. Nokia delivered the first version of the Network Data Warehouse in the first quarter of 1997 and formally published it in Cebit97. Though the functionality of the first version was somewhat limited it received a favourable response from the customers. The impact was most striking in the performance management. A conventional network management system includes a set of extremely flexible tools to discover various aspects of the performance of the network. These tools have required high expertise to use. To overcome this, telecom operators have experienced personnel compiling weekly reports of various key performance indicators they consider important to follow and distributing them to interested persons. The information distribution system cut the time to compile and distribute an actual set of weekly reports from several days to mere minutes and enabled the customization of the information content to various user groups.

Acknowledgements

This paper presents the state of an ongoing project at the Nokia Telecommunications Network Management Systems. Numerous people have made positive contribution to the development of the system, and therefore also this article. The author would specifically like to mention Kimmo Vättö, the soul of the Network Data Warehouse; Jussi Jutila, Mika Hallamaa, Kati Sormunen, Turkka Laakkio and Timo Rasi, the people that finally defined what the distribution system is becoming.

Chinese Information Access Through Internet

Lu Qin, Lee Kin Hong, Yao Jian
Department of Computer Science and Engineering
The Chinese University of Hong Kong
Shatin, Hong Kong
E-mail: {clu,khlee,jyao}@cs.cuhk.hk

1 Introduction

Most of the Internet softwares are designed for English or other alphabet based languages only. Exchange and sharing of Chinese information is very limited. As an ideographic language, Chinese requires a complex processing environment because of the co-existence of multiple codesets, i.e. GB for simplified Chinese, Big5 and CNS for traditional Chinese. The incompatibility of different codesets requires specific supports when Chinese web documents are exchanged via Internet.

One special issue for Chinese information access is the limited local support. Quite a few web browsers can support only one codeset. For example, it is very common that most web browsers on PC platform used in Hong Kong support only Big5 codeset. In such case, it is impossible for users to read documents written in simplified Chinese (gb2312 codeset) with such web browsers. Another special issue is that Chinese documents are expected to be read by people all over the world. People in different regions have different reading preferences, such as people in Mainland China prefer to read documents written in simplified Chinese. To fulfill various reading preferences, most web servers duplicate Chinese documents coded in multiple codesets. However, this method brings out some problems including storage space capacity problem: the more codesets supported, the more space needed, and maintenance problem: inconsistency is liable to happen if only one version has been updated while others have not.

The objectives of our project have three aspects: to provide transparent service for users to access Chinese documents coded in different codesets, to fulfill various reading preferences, but storing Chinese documents in only one codeset, and to provide a customized interface for Chinese users. To realize Chinese information access with codeset transparency, automatic codeset conversion facility should be built in our web system. For example, when people from Mainland China using web browsers supporting only gb2312 codeset want to access a web page in Hong Kong, and since the codeset of the original document(Big5) is incompatible with what users prefer (gb2312), automatic codeset conversion is needed to convert the original document from Big5 to gb2312. In order to handle Chinese text retrieval via WWW with automatic recognition of codesets and the conversion among them, some *codeset announcement* mechanisms must be provided. The client side must be able to announce its local environment or codesets supported. The server side must announce the codeset information for the documents it manages. If the two announcements do not match each other, automatic codeset conversion can then be supported either before the document is transferred or after it is received on the client side. Furthermore, to provide a customized interface for Chinese users, the internationalization concept, I18N for short, is adopted in the design of our browser. It is very easy to switch to different codeset interface with our internationalized browser without modifying the program at all.

In this paper, we describe the design of our web system which consists of a Chinese World Wide Web server, an internationalized browser and an enhanced proxy server [LA94]. Some implementation results are also shown. The Chinese World Wide Web server is built on UNIX platform simiiar to other English Web servers. It can manage Chinese text data encoded in different codesets on the same server and provide automatic codeset conversion transparently to client machines when data stored in the server is incompatible with what the client machine can process. The internationalized browser can work under different languages and cultural conventions. It provides an internationalized user interface. The Chinese Web Browser, a localized version of the internationalized web browser, allows users to access the web server either using traditional Chinese or simplified Chinese without the need to match the server's codeset. The enhanced proxy server is still under development.

2 Architectural Design

The support of automatic codeset conversion is based on *codeset announcement* mechanism which requires codeset negotiation between the client and the server. This can be carried out through the data type negotiation provided by HTTP/1.1 protocol [FFL96]. There are four fields in HTTP/1.1 header: fields `Accept-Charset` and `Accept-Language` in the HTTP request message, fields `Content-Type` with charset parameter and `Content-Language` in the HTTP response message are used for data type negotiation. The first two `Accept` fields provide the codeset announcement mechanism for the web browser. The latter two `Content` fields are filled up by the web server if they follow HTTP/1.1 protocol. With the help of the above four fields, the web browser sends request with the codeset announcement information to the web server. According to this codeset preference information, the web server decides whether an automatic codeset conversion should be done or not, and then it returns the retrieved document to the web browser with the content type information so that the web browser can do proper work accordingly to it.

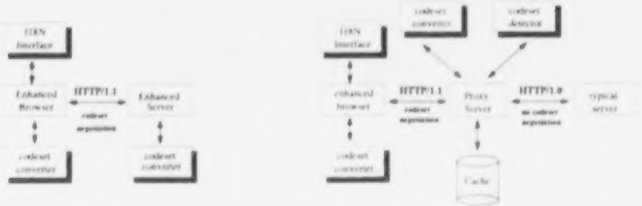
However most of the current web browsers and web servers support pre-HTTP/1.1, such as HTTP/1.0 [LFF95]. Besides providing an enhanced server and an enhanced browser supporting HTTP/1.1, the system should be compatible with web systems which support pre-HTTP/1.1 protocol. This brings our design requirements of flexibility for backward compatibility. The component integration approach is used in our architectural design, where each component is independent and reusable, and all components can act in flexible combination to provide services under different situations. There are four cases for the framework of our system which are illustrated in the following figures. All rectangles with shadows are characteristic modules in our web system. Those drawn with dashed lines are optional modules or functions.

Case I: Enhanced Browser Accesses Enhanced Server

Figure 1(a). shows that an enhanced browser can communicate with an enhanced server without any problem. The enhanced browser provides an internationalized interface for users, and both the server and browser carry out data type negotiation according to HTTP/1.1 protocol.

Case II: Enhanced Browser Accesses Typical Server

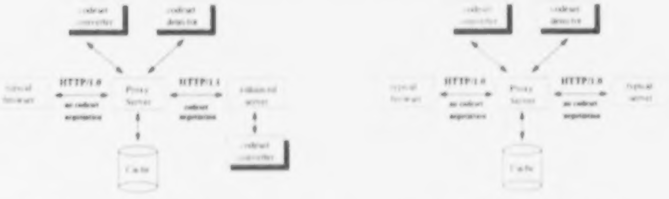
Figure 1 (b). illustrates if an enhanced browser wants to access documents on a typical server. Since the typical server will ignore the codeset announcement information sent by the enhanced browser, and the typical server doesn't have any codeset conversion facility built in, an enhanced


Interface
Enhanced Browser
HTTP/1.1
codeset negotiation
Enhanced Server
codeset converter
codeset converter
Interface
codeset converter
codeset detector
enhanced browser
HTTP/1.1
codeset negotiation
Proxy Server
HTTP/1.0
no codeset negotiation
typical server
codeset converter
Cache

(a) Enhanced Browser Accesses Enhanced Server

(b) Enhanced Browser Accesses Typical Server

Figure 1: Case I and Case II


codeset converter
codeset detector
typical browser
HTTP/1.0
no codeset negotiation
Proxy Server
HTTP/1.1
codeset negotiation
enhanced server
codeset converter
Cache
codeset converter
codeset detector
typical browser
HTTP/1.0
no codeset negotiation
Proxy Server
HTTP/1.0
no codeset negotiation
typical server
Cache

(a) Typical Browser Accesses Enhanced Server

(b) Typical Browser Accesses Typical Server

Figure 2: Case III and Case IV

proxy server must be added as a bridge between them. In this case, the proxy server accepts the request from the enhanced browser, and forwards it to the typical server. After receiving retrieved document from the server, the proxy server will try to identify the codeset of the retrieved document and do automatic codeset conversion if necessary. Then it returns the converted document to the browser.

Case III: Typical Browser Accesses Enhanced Server

Figure 2 (a). shows another case where a typical browser wants to access an enhanced server. Although the enhanced server can do codeset conversion, however, the typical browser doesn't send any codeset announcement information to the server, there is no way for the server to know which codeset it should convert the document to. As a result, an enhanced proxy server is also needed in this case. In general, a proxy server is located within a local network. The assumption used by such proxy server is that it regards that most of the web browsers within the local network support an identical codeset. For instance, most web browsers in Hong Kong support Big5, so that the proxy server can announce this codeset to the enhanced server and automatic codeset conversion can be done based on this information.

Case IV: Typical Browser Accesses Typical Server

If users don't have our web software at hands, another framework is also provided for them, which is shown in Figure 2 (b). In this case, an enhanced proxy server is a must if users still want to have the specific services. The proxy server assumes that a particular codeset is supported by most web browsers in the local network, and then it accesses the document on the server on behalf of these web browsers. It is indispensable to enhance the proxy server with automatic codeset detection facility so that it can identify the original codeset of the retrieved document by investigating the source code of the document, and carry out codeset conversion if needed.

The enhanced proxy server with caching capability also speeds up the document retrieval when a web browser accesses the document which has been accessed by another browser before. The proxy will fetch it from the cache instead of connecting with the remote server again.

BEST COPY AVAILABLE

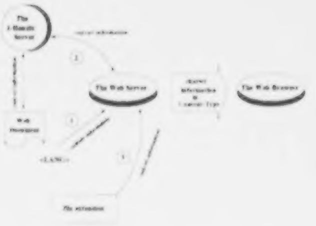


Figure 3: Steps in Determining the Codeset of a Given File

3 Design Details and Implementation

The whole system consists of three parts: an enhanced server, an internationalized browser and an enhanced proxy server.

3.1 The enhanced server

The functionalities which an enhanced server should have are as follows:

1. **Interpretation of Client's Codeset Announcement:** Analyze field *Accept-Charset* and field *Accept-Language* in HTTP/1.1 request message. These two fields are used by the client to announce the codeset it supports or prefers.
2. **Codeset Identification of Web Documents on the Server:** To carry out codeset conversion, the prerequisite is to know the original codeset of the retrieved documents on the server. The language or codeset tagging of a web document can be done through the tag <LANG> defined in HTML/3.0 [Rag95]. With this <LANG> tag, it is easy to compose a single document with multiple codeset parts. Other two methods used by our server is to utilize our I-Hanzix server [han94] developed before to tell the codeset of the document, or using file extension convention, i.e. file extension .htmlgb means the document is written in gb2312 while file extension .htmlb5 means the document is written in Big5. The procedure of codeset identification is shown in Figure 3.
3. **Codeset Conversion:** If the original codeset of the retrieved document is incompatible with the codeset the client supports or prefers, codeset conversion should be carried out to convert the retrieved document from the source codeset to the target one.
4. **Codeset Notification to the Web Client:** Whether codeset conversion has been done or not, the server should notify the codeset of the retrieved document to the client. This information can be contained in the field *Content-Type* with codeset parameter. Based on this notification, the client can do further codeset conversion if necessary or just display the document in a proper environment.

3.2 Internationalized Browser

The enhanced browser must be able to announce its codeset preference information when sending out a request to the server, and it is also capable to interpret the codeset notification information contained in the response message. Codeset conversion facility should also be incorporated in the enhanced browser. To provide a familiar operation environment for users, it is necessary to customize the browser interface to support different codesets. Instead of using the traditional remove-replace or add-on approach, internationalization concept (I18N for short) [xop90] is adopted in the development of our enhanced browser. This new concept isolates the program context data related to language or cultural conventions from the code itself. All program context data, such as prompt information, help or error messages, menubars or buttons are saved in

BEST COPY AVAILABLE

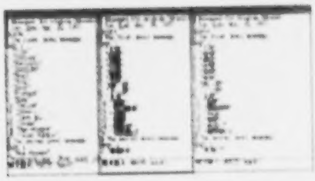


Figure 4: The Fragment of The Message Source File

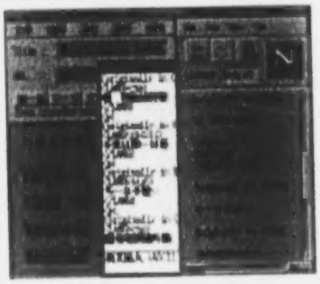


Figure 5: A single file written in multiple codesets

message catalog system [O'D94] separately. The message catalog system usually contains several versions of the program context data for different languages or codesets. Figure 4. shows the fragment of the message source files in our message catalog system. In our system, three codesets are supported currently: gb2312, Big5 and CNS11643. Unicode will be added into our system in the next step.

Currently, we have finished the development of our enhanced server and our enhanced browser. The server is developed based on CERN's httpd 3.0, the browser is based on Mosaic-2.6. Both of them are free software for Unix platform in the public domain. The enhanced proxy server is still under development. Figure 5. shows that a single document written in three codesets simultaneously (gb2312, Big5 and CNS 11643) using tag <LANG> can be converted to the same target codeset - gb2312 using our web system, while Netscape 3.0 can display only the lines compatible with the current encoding setting. Figure 6. illustrates that our server can identify the original codeset of the retrieved document *vr.htmlgb* and carries out codeset conversion from gb2312 to Big5. Figure 7. shows the customized browser interface for two codesets: gb2312 and Big5. The programs for these two versions are the same, but the environment settings are different.

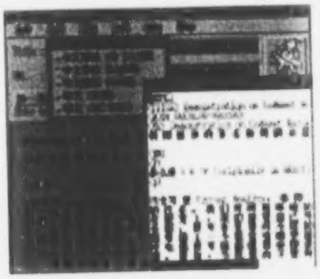


Figure 6: Codeset Identification Based on File Extension

BEST COPY AVAILABLE

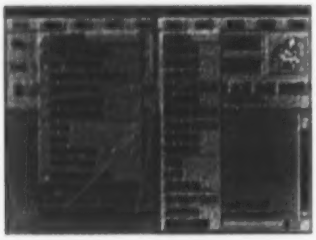


Figure 7: Customized Browser Interface Under Two Codesets

4 Conclusions and Future Plans

The development of a web system providing Chinese information access via WWW with codeset transparency is very useful for people in Mainland China, Hong Kong, Taiwan and others who read and use Chinese. The component integration design is very flexible to fulfill various requirements with high efficiency and backward compatibility through the optimal combination of related components. Currently, a Chinese web server and an internationalized web browser has been developed and they can communicate with each other through data type negotiation defined in HTTP/1.1 protocol. The enhanced proxy server is still under development.

The component integration concept is applicable to both Unix platform and PC platform. The next step of our project is to develop such web system for PC platform. As Unicode becomes more and more important and popular, it is necessary for us to support Unicode and codeset conversion between Unicode and other Chinese codesets in the future. Furthermore, the approach can also apply to document access between different languages. The codeset converter can be replaced by an intelligent language translator. In this way, document written in Japanese may become accessable to users who doesn't know Japanese.

References

[FFL96] R. Fielding, H. Frystyk, and T. Berners Lee. Hypertext transfer protocol http/1.1 (http working group, internet draft), January 1996.

[han94] The hanzix open system. In *Proceedings of International Conference on Chinese Computing '94*, Singapore, June 1994. Hanzix Work Group.

[LA94] Ari Luotonen and Kevin Altis. World-wide web proxies, April 1994.

[LFF95] T. Berners Lee, R. Fielding, and H. Frystyk. Hypertext transfer protocol http/1.0 (http working group, internet draft), September 1995.

[O'D94] Sandra Martin O'Donnell. *Programming for the world - A guide to internationalization*. Prentice-Hall, 1994.

[Rag95] Dave Raggett. Hypertext markup language specification version 3.0 (http working group, internet draft), March 1995.

[xop90] *Internationalization Guide*. X/Open Company Limited, 1990.

BEST COPY AVAILABLE

VIRTUAL WORLDS OF TODAY, VIRTUAL WORLDS OF TOMORROW

Howard Rosenberg, Rosenberg & Associates, USA (e-mail:howie@ix.netcom.com)

INTRODUCTION

VRML (Virtual Reality Modeling Language), though still in its' infancy, is rapidly changing the Internet landscape into a 3-dimensional medium. Developments in VRML and JAVA, are opening new and exciting vistas to all levels of the educa- tional community. Exploration of these environments, is an endeavor which may have value far beyond simple edutainment. On-line social interactions, developed while navigating through 3-dimensional spaces (ie. SonyLabs Community Space; http://vs.spiw.com/vs/), may lead students to work on colloborative projects in various areas of study (gaining useful technical skills along the way). Navigational skills learned while exploring these environments, may spur students to explore career paths, such as astronomy and aerospace. The first section of this paper will discuss several virtual worlds which have already been created and deployed over the WWW. The second segment will focus on several virtual environments which could be created in the future.

VIRTUAL WORLDS OF TODAY

I. VIRTUAL LOS ANGELES (http://www.planet9.com)

Logging onto the Planet9 web site, viewers have several fascinating 3-dimensional worlds and cities worthy of explora tion. Students may travel along the highways of Los Angeles using the navigational tools on the bottom of the display. The 3-dimensional polygon may be rotated("spin" & "slide" features), navigated ("walk" & "look" features) and panned in and out of ("point" feature). The lighting of the site may, likewise, be changed by using the "lamp" feature. This particular site is unique, in that it contains one virtual environment within another. As one reaches Beverly Hills, while exploring the Los Angeles area, one may hyperlink to a virtual display of Beverly Hills. This virtual environment appears in the adjacent frame, with a virtual Rodeo Drive that can be explored.

II. VIRTUAL OPERA -- SAN FRANCISCO (http://www.planet9.com)

Observing the virtual opera of San Francisco appear on screen, is akin to watching a work of art take shape before one's eyes. Though downloading may at times be a lengthy and arduous process, watching rows, balcony and stage appear, gives viewers a genuine feeling of being at the opera. Equipped with an avatar of a singer on stage, the experience will become even more realistic as audio files of classical opera pieces are added to the site. As in all the virtual environments mentioned, if one becomes lost or disoriented while rotating or navigating within the display, clicking on the "view" button, will always return the display to its' original position.

III. VIRTUAL SOLAR SYSTEM (http://www.planet9.com)

Exploration of the galaxies, has long been an area of study, fascinating to both students and educators. This site allows viewers to explore and probe the depths of the universe. As one takes the navigational controls, planets, their moons, comets and the Sun are just several of the celestial bodies that fill the screen. Celestial bodies moving by, as one navigates deeper into space, create an effect where the traveler feels like he/she is traveling through space. This is a trip worth taking, by all cyberexplorers.

VIRTUAL WORLDS OF TOMORROW

VRML is providing Internet content creators with the tools necessary for stretching the boundaries of the medium. This segment of the paper will propose several 3-dimensional worlds that could be constructed, that would allow students to explore environments in an entirely different manner, enhancing their educational experiences tremendously. Each of the proposed environments, may be used as a springboard for creating other 3-D worlds of a similar genre. Bandwidth limitations and time constraints, would of course, need to be taken into consideration when incorporating these displays into classroom curricula.

I. EUROPEAN COMMUNITIES OF THE 20TH CENTURY

The European landscape has undergone seismic changes during this century. Two World Wars and umpteen regional conflicts have decimated many European communities. VRML offers the educational community the opportunity of recreating these lost communities on-line. The use of avatars could allow students the opportunity for interactions with people in environments, long lost to the annals of history. Used as a supplement to traditional pedagogical approaches to the study of European history, classes could be designed so that students could explore various "what if" historical scenarios. Through this process, they may learn how they would cope with various crises and challenges. Linking these virtual environments to digital libraries and museums, could likewise enhance researchers' efforts.

II. VIRTUAL TREK THRU THE HIMALAYAS

Only a fraction of the planet's inhabitants, will get the opportunity to gaze upon or explore the Himalayan mountain range. It is for the majority, that a 3-dimensional tour of the Himalayas would be a fascinating educational experience. Such a virtual exploration could embellish on areas of study such as geology, earth science and geography. Simulating the Himalayan treking experience, may lead to development of an interest in the study of other mountainous regions and the Tibetan culture.

III. ATLANTIS RECLAIMED

Long a source of mystery to historians, archeologists and explorers, the lost city of Atlantis can be recreated using VRML, allowing viewers to explore the underwater depths in search of the city. Such a tour could be incorporated into a class on marine biology, with the possibility of hyperlinking to databases of information on marine biology. Digitally recreating Atlantis could be one of many Greek mythologies which could be created using VRML, thus broadening the educational horizons of students.

A Dynamic Visualization of Dialog Contents in a Collaborative Communication

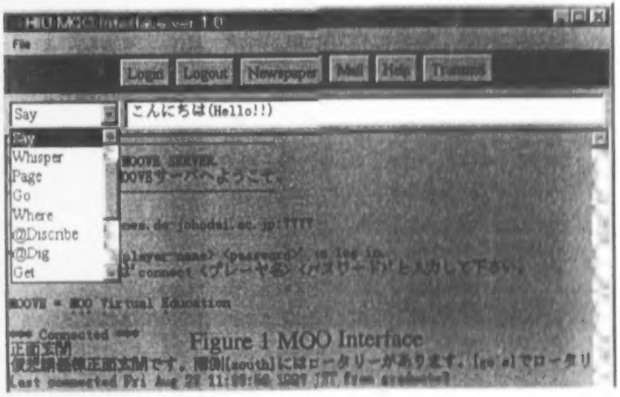
Hajime SAITOH Takashi MAEDA
Graduate School of Hokkaido Information University
59-2, Nishinopporo, Ebetsu, 069 JAPAN
E-Mail: {hajime, maeda}@do-johodai.ac.jp

1. Introduction

People might sometimes get a new idea, a clue or a hint for some problem-solving as well as information exchange through various conversations or discussions. Such conversations or discussions have recently been ubiquitous not only in the real world such as at home or in the laboratory, but also in some virtual communication environments such as computer networks like the Internet. These communications are usually transient in nature, but messages like electronic mail can be referenced and reviewed again, if saved, though these processes are one-way only and cannot be interactive among the participants in that communication.

We investigate a method here to represent effectively the key contents of dialogs for more rich communication and creative thinking, etc. In this paper, by using a visualization technique, we describe a method that will extract, sum up and display key contents of dialogs in natural language, currently in Japanese as a conceptual map, by several participants in the communication.

2. Collaborative communication environment



こんにちは(Hello!!)

MOOVE SERVER

MOOVE = MOO Virtual Education

*** Connected ***

Figure 1 MOO Interface

The MOO[1] system we use in this paper is a type of the MUD of object-oriented, and it has been widely used as a social and interactive role-playing system on the Internet. The participants in the collaborative communication on the MOO maybe behaved as a "character" supported by client program via telnet. They can virtually communicate with each other through text-based messages. The communication environment can be extended to various applications, for example, referring the same WWW page including the same link among the participants by combining established WWW.

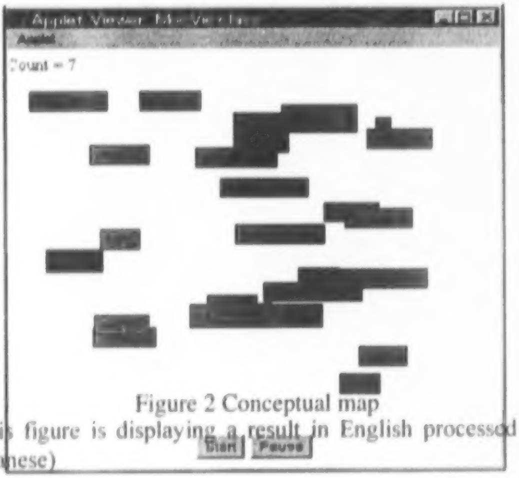


Figure 2 Conceptual map
(This figure is displaying a result in English processed by Japanese)

We have a plan to add more supporting functions that will enhance collaborative communication in addition to the visualization of dialog contents, for the MOO system itself is still basically text-based[2].

3. Visualization method

We have endeavored to use keywords in dialog text as a representation of key contents of dialog , then to arrange these keywords in a two dimensional map according to the relationship among them. The relationship among keywords has currently been calculated using "the spring model[3]" based on the dynamical interaction.

Our visualization method consists of three main process:

Gathering dialog contents which is text-based
Summing up keywords in the contents
Calculating and displaying the

BEST COPY AVAILABLE

relationships of these keywords using the spring model

An interface developed to easily access the MOO environment includes a function to gather and save their messages(Figure 1). Next keywords in the messages are filtered and located on the two-dimensional space along the above processes. Thus, we can dynamically display these maps, called conceptual maps, by calculating these relationships step by step and represent interactively these relationships among the keywords in response to the progress of dialogs.

Figure 2 is an example of a visualized map of the contents of a dialog about MOO itself by our method.

4. Results and Discussion

Our method at this stage could be evaluated at least by the following two points; first, we have constructed a dynamic visualization method of dialogs in spite of fragmentary information sources, and second, the method can emphatically display keywords to be used frequently as more important ones. As a result, we believe that this method can be available to support more understanding of communication.

We have used the spring model to extract and display some of the relationships of keywords in rather static information(ex. Netnews , E-mail), so that it has some limits to realize a communication supporting function. First, keyword location can't be effected in the map in real time, because the relationship among keywords in the fragment is calculated independently. Second, each client can not possible perform interactive operation and edition in these processes. Finally, we use only a part of the abilities of the natural language analysis system for Japanese, "Chasen[4]" , so that any contextual information in the dialog is not considered in the process.

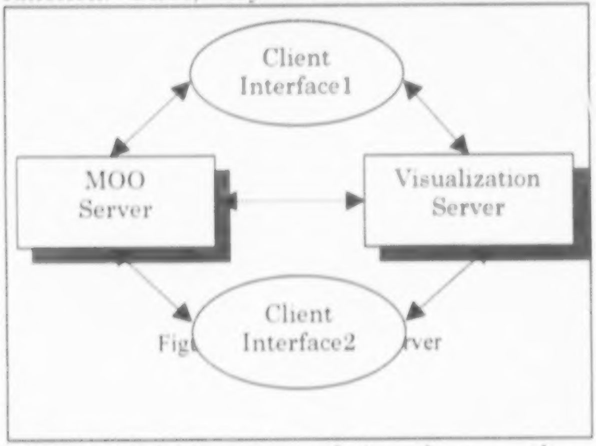


Fig ver

We plan to refine this method as follows(Figure 3). First, we will adopt a client-server construction to separate some functions of the method. Second, natural language analysis will be performed in each visualization server, so that various representations can interactively be displayed on each server site in real time. Thus, participants will be able to concurrently handle such a representation of the total contents of dialogs and of the messages of all participants, respectively.

5.Conclusion

In this paper, we have described a method of visualization of dialogs to enhance communication and understanding. Based on this method, we have pointed out future plans including refinements of this method. From now on, we will aim at improvement of these points.

Acknowledgement. We would like to thank Dr.P.Dillenbourg of TECFA, University of Geneve for introduction and some suggestions on the MOO system as a virtual communication environments in computer networks and Charles McLarty for his comments on our manuscript.

References

[1] S Tognotti, D K Schneider, P Mendelsohn, Analysis of MOO and WOO Environments, TECFA,Universite de Geneve, 1995

[2] Takashi MAEDA, Hajime SAITO, Takehisa YANAGAWA, On Constructiong of a Collaborative Communication Environment,MEMOIRS of Hokkaido Information University 1996

[3] K TAKASUGI, S KUNIFUJI, An Idea Inspiration System using Spring Model, Journal of Japanese Society for Artifical Intelligence,1996

[4] Y Matsumoto, et al. Japanese Morphological Analysis System ChaSen Manual, Nara Institute of Science and Technology 1997

Designing a Context-Specific Questionnaire for Online Teaching

*Denise Stockley & Vivian Rossner
Simon Fraser University,
Burnaby, British Columbia, V5A 1S6
dstockle@sfu.ca; vivian_rossner@sfu.ca

* This paper was written collaboratively and order of authorship is not reflective of contribution.

This work-in-progress will provide the participants with a description of the Simon Fraser University's LohnLab, a description of the development of a context-specific questionnaire for the lab, the results of pilot testing and future directions for the questionnaire based on the pilot results. The LohnLab is a unit located within SFU's Centre for Distance Education and is a member of the University's Instructional Development Group. The Instructional Development Group is a unit that combines the expertise and resources available from the Library, the Centre for University Teaching, the Instructional Media Centre, and Academic Computing Services. The Lab provides professional and support staff to ensure that a pedagogical and technological blend of competencies is fully available to an anticipated range of faculty expertise. One of the roles of the support staff is to provide leadership and training in using a variety of approaches to Web-based teaching and learning (e.g., tele-apprenticeship, reciprocal teaching, collaborative learning, peer interaction, role playing, simulation, information access, or other pedagogical models suitable for the Web an instructor may wish to use).

The LohnLab team is committed to provide pedagogical and technological support that is based on the interests and needs of the faculty members using the lab. To help ensure this goal was being met a questionnaire was suggested to the lab team. the initial plan was to use a "ready-made" computer attitude questionnaire (e.g., Francis & Evans, 1995; Kay, 1989). After a comprehensive literature review and web search we found that yes there were many neat, reliable and valid questionnaires available but, they addressed attitude-type questions that did not address our context-specific issues. For example, we wanted to know what the faculty thought of the lab as a pedagogical and technological support centre, how they planning on using the lab, and how the lab could accommodate their needs more easily. To address these issues a questionnaire was designed specifically to be used in the LohnLab setting. This questionnaire is both qualitative and quantitative in nature and includes demographics, some computer attitude questions and context-specific questions. After many drafts and discussions a pilot-ready version was created.

The questionnaire is divided into four sections. The first section on learner characteristics provides demographic and background information about the users of the lab. This information will provide the lab staff with contact information such as e-mail, phone number, and whether they are a continuing faculty member (a potential long-term client) or a graduate student who will use the Lab minimally. The second is based on course specific questions which change the direction of the questionnaire from general background information to their purpose in using the LohnLab as a resource for online teaching. These questions focus on what they are interested in teaching and whether they choose the course or were assigned to teach it (this may affect the instructor's willingness to learn the technology). The third section is a series of Likert questions that target instructor attitudes towards a number of items related to online teaching. Attitudes is defined as relatively stable orientations toward a part of an environment and have three distinct areas. These include: (1) a behaviour component which reflects an individual's action toward the computer; (2) an affective component which indicates a person's inward feelings toward the computer; (3) a cognitive component which is the belief held about computers (Brown, Brown, & Baack, 1988). These questions are on a 5-point Likert scale ranging from 1=strongly disagree to 5=strongly agree. The questions will be checked for reliability and validity through pilot testing. The questions target a variety of attitudes, including technology issues, learning online, teaching online, pedagogical awareness and the LohnLab. The final section of the questionnaire is a series of open-ended questions. These questions target perceptions of online teaching, planning, attitudes and needs of the instructors.

This questionnaire is currently being piloted to the larger community of LohnLab users and preliminary results will be presented at WebNet.

References

Brown, T.S., Brown, J.T., & Baack, S.A. (1988). A reexamination of the attitudes toward computer usage scale. Educational and Psychological Measurement, 48, 835-842.

Francis, L.J., & Evans, T.E. (1995). The reliability and validity of the Bath County computer attitude scale. Journal of Educational Computing Research, 12(2), 135-146.

Kay, R.H. (1989). A practical and theoretical approach to assessing computer attitudes: the computer attitude measure (CAM). Journal of Research on Computing in Education, 21, 456-463.

A web based automated advisor for delivering purchasing advice

Sandor Szego
The Institute for the Learning Sciences
Northwestern University, U.S.A.
szego@ils.nwu.edu

Introduction

The number of commercial websites has increased dramatically during the past few years, and this trend is certain to continue as consumers become more familiar with the web and as security issues are gradually resolved. Web technology has also seen many advances: scripting languages, newer versions of HTML, Java, just to mention a few. However, in the midst of all these changes and improvements, one thing has remained the same or became even worse: individual websites and the web as a whole are very poorly organized.

A special kind of search engines, on-line yellow pages servers, are currently the only web based applications that help users find the right product(s) for their needs. The interaction with such engines is relatively simple. The user enters a product/service category name, and the search engine returns a list of websites that are sellers or providers of the selected product or service. (Of course sometimes the query returns nothing, and the user has to guess at the product/service category name again until the right category is found.) After the category is found, the user can visit and navigate through each site in turn, independent of the others listed by the search engine.

Our goal in this research is to develop a new brand of retrieval and information organization systems for commercial websites to better support the task of purchasing. There are three features that differentiate our proposed new system from the standard yellow pages model. First, finding products or services is based on zooming (Osgood, 1994) instead of querying. Zooming obviates the need for guessing, because at any point of the zooming process the user is presented with a set of meaningful choices that can take the user closer to the product/service he or she is interested in. Second, instead of providing users with the "homepage" of a target website[1] , the system provides a view of the website that is tailored to the users' needs. Finally, once users find the product or service of their interest, the system presents a browsing interface through which potentially relevant information from other websites is made available (c.f. ASK systems, in Osgood, 1994).

The Shopper's Assistant Browsing System

The interaction with the Shopper's Assistant consists of two phases, zooming and browsing. In the zooming phase, users find a *product* or *service* they need or an *activity* that is the motivation for trying to purchase something. For instance, if a person has decided to buy a tent because he/she wants to go hiking, then either the product category "tents" or the activity category "hiking" are possible outcomes of the zooming phase.

The system provides two graphical zooming interfaces. One is based on locations: it displays prototypical locations using icons on a cartoon-like map. (A similar interface is described in [Domeshek et. al 1996].) The idea behind this zoomer is that users can readily identify the location where a product/service is typically used or needed (see Schank & Abelson, 1977 for an elaboration on this idea).

---

[1] The phrase "target website" refers to a website that contains relevant information for the user (e.g. a website that sells a product or service, or that contains information about products and other commercial websites). In a traditional Yellow Pages application a "target website" simply means the website of the seller or service provider.

By clicking on a location, users can either get a close-up view of the location, or they can get a list of products/services/activities that are associated with that location. For example, if we are looking for the product category "tent", mountains or forests are a good place to start to search for this category, since tents are often used during hiking in the mountains. (Not surprisingly, the activity category "hiking" is also found there.) The other zooming interface is the "life span zoomer", a graphical representation of major events in a person's life (such as birth, schooling, weddings, retirement, etc.). This zoomer is best to use when the purchasing occasion is related to one of these major events.

After the zooming step, users are presented with a browsing interface. In the center of the browser is a product information box from a vendor of the user's choice. This information is generated on the fly based on data received from the vendor. Around this area is a standard set of browsing buttons. These buttons can be used to retrieve additional information on the selected product, or they can be used to retrieve sellers of related products and services. The browsing relations fall in the following categories: vendor specific (e.g., "About this vendor" button retrieves information about the current vendor), related products (e.g. "Accessories" retrieves vendors which sell accessories for the selected product), related services (e.g., "Rental" retrieves renters of the selected product), reviews (this button retrieves sites providing reviews on the selected product), and related activities which retrieves information providers about activities related to the selected product. By using these browsing relations not only can the user get information that is directly relevant for making the current purchasing decision, but these relations can also suggest other products the user might need (e.g. a "footprint" to protect the floor of the tent), or alternative ways to achieve the user's goal (e.g., renting instead of buying a tent).

How the Shopper's Assistant works

The key component of the Shopper's Assistant is a richly interconnected set of product, activity and service categories. Activity categories form the backbone of the network, while products and services are connected to these activities through different links. Using these basic links the system can infer other relations. For instance, the product categories "backpacks" and "tents" are both linked to the activity category "hiking" through the "used-in" link. This allows the system to infer that "backpacks" and "tents" are related products, because they are used in the same activity. Sites are then indexed using a category (or a set of categories if the site sells more than one product category) and a modifier. The modifier indicates whether the site is a seller of the product or it is a product review provider.

Another key feature of the Shopper's Assistant is that it displays product specific information from a vendor when the user focuses the browser on a product. However, this product specific information is not stored on the Shopper's Assistant server, instead it is requested from the actual vendor when the information is needed. Therefore, this feature may require extra effort from the developers of websites if they want to be registered with the Shopper's Assistant server. However, our intention is to create website authoring tools that would not only make the development of commercial websites easier, but they would automatically provide the information needed by the Shopper's Assistant (or similar servers).

Conclusions

This paper is based on the work in progress to provide intelligent interfaces to commercial websites. We are currently finishing the prototype of the Shopper's Assistant server, and we are in the process of designing content rich website authoring tools to complement the Shopper's Assistant system.

References

[Domeshek et al., 1996] Domeshek, E. A., Kedar, S. & Gordon, A. (1996). Interactive Information Retrieval Systems with Minimalist Representation *AAAI-96: Thirteenth National Conference on Artificial Intelligence*, Portland, OR.

[Osgood, 1994] Osgood, R. E. (1994). The Conceptual Indexing of Conversational Hypertext. Ph.D. thesis, Northwestern University.

[Schank & Abelson, 1977]Schank, R.C. & Abelson, B. (1977): *Scripts, Plans, Goals and Understanding*. Hillsdale, NJ: Lawrence Erlbaum.

Collaborative Tasks and Outcomes in Online Training: The Infoshare Module

Lucio Teles and Xinchun Wang
Simon Fraser University

This paper discusses the design, development and implementation of a collaborative learning environment for online training. The module "Infoshare on the Web" was designed to teach participants how to use Web search engines. Participants collaborate in completing group tasks using asynchronous communication provided by Simon Fraser's Virtual-U, a Web-based environment that supports distance education and information sharing.

Collaborative Online Asynchronous Training

> It has been noted that students no longer go to universities just to acquire a finite body of knowledge; they now want to "learn how to learn," how to renew themselves continuously intellectually in order to keep pace with the demands that will be placed on the knowledge worker of the twenty first century. It is predicted that for a person to remain gainfully employed in the emerging knowledge economy an equivalent of 30 credit hours will be required every seven years. (Horvath & Teles, 1997).

The main product of the knowledge economy of the next century will be knowledge itself. Workers will need ongoing education to keep up with the demands of society, and online learning provides another mode for obtaining it (Harasim, Hiltz, Teles, Turoff, 1995; Hiltz, 1994).

Online classrooms are already being used to offer credit courses and to support information sharing, decision-making, and collaborative tasks. They facilitate knowledge sharing and are increasingly being used in many subject areas (Richards et al, 1997). They can be customized to reflect an instructor's own approach to teaching, and they can provide a number of special tools to assist the learner. Recently, online environments have also been customized for corporate training (Harasim, Hiltz, Teles, Turoff, 1995).

Online asynchronous training allows participants to have access to information and engage in collaborative tasks and ongoing discussions at times that fit their own schedules. Collaborative learning may allow participants to have multiple perspectives on a topic (Harasim, 1997; Langer, 1997), which fosters the development of their problem-solving skills as they undertake particular tasks. Synchronous tools can also be used to support participants' skills development.

Research Questions

Three research questions are addressed in this paper:

1. Can trainees acquire or improve Internet-related skills through collaborative asynchronous tasks in online classrooms?

2. What are the patterns of collaboration as indicated by participants' messages?

3. What type of messages do participants use to communicate: text only, text + hyperlinks, text + multimedia, text + hyperlinks + multimedia?

The InfoShare Module

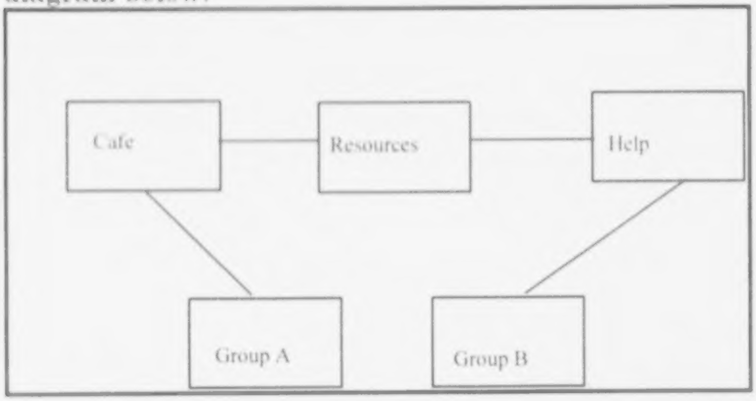
The Infoshare Module was designed to teach participants how to use (or to improve their use of) Web search engines to find information and share it with others and to test the use of online environments to support this process. The course was offered from mid-September to mid-October 1996.

The 24 active participants were professionals in a national research institution based in British Columbia, Canada. They advise Canadian companies on various matters, including patents, intellectual property, marketing, and global exports.

The participants, who were located in various cities, were split into two online groups of 12 members each. The Infoshare module was delivered over a three-week period, with each week containing a topic and a task. In the introductory face-to-face session participants were trained in the use of Virtual-U and provided with course material, information, and module requirements. The following three sessions were entirely online and focused on introducing Web search engines to participants and having them generate a Web resource list containing sites relevant to patents, intellectual property, marketing, and financial planning.

Course Design

Each participant had access to four conferences, as shown in the diagram below:



The Cafe was for informal discussion, Resources, for sharing information, Help, for technical help, and either Group A or B, for collaborative tasks. Participants also had a group email list. They could reach the instructor via one-to-one email or conference messages.

At the introductory face-to-face session articipants were given print material containing information about course objectives and topics for each session, which was also available online. Additional readings included online articles from other sites, which could be reached by pointers.

The course sessions began on Monday mornings with an electronic lecture followed by individual and group tasks. Since the objective was to learn how to use the search engine Alta Vista, participants were given tasks that required the use of instructions in Alta Vista's Manual for Short and Advanced Commands. For group tasks, participants used the two group conferences. In a typical module task (week three), participants had to identify a Web site using an Internet search engine (preferably Alta Vista). In the fourth week, they were expected to generate a list of Web sites relevant to their work.

Methodology

The data sources for investigating collaborative tasks in online environments and participants' interaction patterns are transcript analysis of conference messages supported by usage statistics Transcript analysis was also used to assess whether or not the module's goals were attained. Conference messages were analyzed to identify what type of messages they used to communicate (text only, text + hyperlinks, text + multimedia, text + hyperlinks + multimedia) and to identify communication patterns.

Results

While there were hardware problems that affected the project, for example, by slowing down the networking connection, a total of 228 messages were generated, an average of 9.5 messages per participant.

The following table summarizes the interaction patterns of one conference. The category "Participants' responses to course topic" represents messages that were students' responses to the weekly topic. "Participant-to-participant messaging" refers to messages participants sent to one another. "Instructor's topic/assignment" refers either to instructor's introduction to the topic of the weekly class session or to tasks participants were expected to conduct. "Instructor's responses to participants" are instructor's replies to participants' questions or comments on participants' tasks.

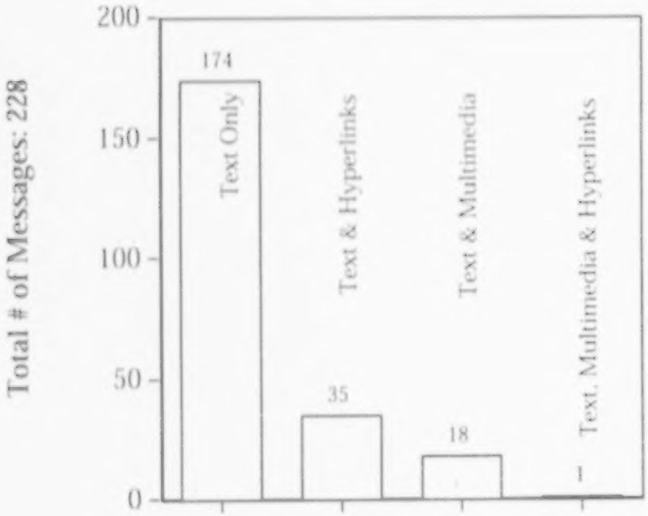
Interaction patterns	Number of Messages
Participants' responses to course topic	26
Particip-to-particip messaging	36
Instructor's topic/assignment	7
Instructor's responses to participants	12
Total	**81**

Most messages were of the collaborative type with participants asking a question, addressing a colleague, or responding to someone's question or comment. The instructor produced 19 out of 81 messages.

Conclusions

The total number of messages generated in three weeks is significant and shows active participation. In some cases a particular topic of discussion could lead to a thread of discussion containing three or four messages.

Most of the messages written in the module were peer to peer messages, which shows a high level of interaction and collaboration. As shown in Message Type chart, most of the messages were text only, but hyperlinks were frequently used, and some multimedia effects were added to messages, such as pictures, diagrams, etc.


Message Type

Participants were able to develop sufficient search skills in the online classroom to generate a list of 33 Web sites relevant to their work.

More research is still needed to investigate the system features that best support collaborative work and to determine which training techniques best work in asynchronous environments.

Bibliography

Harasim, L., Hiltz, R., Teles, L., & Turoff, M. (1995). *Learning Networks: A Field Guide to Teaching and Learning Online.* Cambridge: MIT Press.

Harasim, L., Calvert, T. and Groeneboer, C. (1997). Virtual-U: A Web Based System to Support Collaborative Learning. In B. Khan's, *Web-Based Instruction*, Englewood Cliffs, NJ, Educational Technology Publications.

Hiltz, S. R. (1994). *The virtual classroom: A new option for learning via computer networking.* Norwood, NJ: Ablex.

Langer, E. (1997) *The Power of Mindful Learning.* New York: Addison-Wesley Publishing.

Lick, D. (1996, April). Ivory Towers and Ivy-Covered Walls will Yield to Online Learning. *The Christian Science Monitor.*

Richards, S., Barker, P., Tan, C., Hudson, S. and Beacham, N. (1997, February) Knowledge Sharing Through Electronic Course Delivery. pp. 3-10. *Innovations in Education and Training International,* Volume 34, Number 1

Computer-Mediated Communication in Enhancing Communicative Dialogues on the Web

Seppo Tella & Marja Mononen-Aaltonen
Media Education Centre, Department of Teacher Education,
University of Helsinki, Finland
(seppo.tella@helsinki.fi) (mononena@cc.helsinki.fi)

A Three-Dimensional Model of Communication Channels

Our presentation aims at describing a conceptual framework focusing on the three-dimensional model of communication channels (monologic; dialogic; telelogic). The model will be analysed through HHC (human-to-human communication) vs. CMC (computer-mediated communication). A special emphasis will be put on computer-supported collaborative work (CSCW). Our primary goals include the cultivation of the students' expertise when enhanced through a carefully reflective scaffolding of the teacher's own command of computer-mediated human communication.

The Impact of ICT on Human Communication

The modern information and communication technologies (ICT) have a two-fold impact on human communication: (i) they open up new potential to increase mutual understanding worldwide and (ii) they provide teachers and students with new tools and techniques for human-to-human communication (HHC). Human-to-human communication is intrinsically culture-based and as such always conveys the threat of misunderstanding and intercultural clashes.

Different Contexts of Dialogues

In our research project, the term 'dialogue' is used in three different contexts. First, our analysis of human-to-human communication (HHC) and computer-mediated communication (CMC) will be based on the notion of dialogue as the basis of all human communication and interaction, and we will extend it to understanding different cultures. Second, dialogue is the key concept in the teaching/learning process. Third, we use dialogue as a general definition of indivisible origins of thinking. Our main argument is that dialogue is becoming a crucial element in the creation of any learning organisation and especially in establishing an open multimedia-based collaborative and networked learning environment. In our research project, we will lay special emphasis on enhancing the students' capacity to communicate in a network-based learning environment provided by the WWW and other telematic tools.

Communicative dialogues make us understand that most of what is significant to human beings is in one way or another created through shared talk and negotiated meanings, and that there is enormous transformative power in this activity as its nature and impact are gradually understood. Deeply connected to this is the recognition of the fact that new dialogic levels can produce new levels of coordinated action, especially when working on the Web and equally between human beings.

The Dominance of Voices

One of the dimensions in the implementation of ICT is the dominance of voices. Tella [1997] has illustrated how the three stages of this dimension--monophony; stereophony, and polyphony--are related to the development of ICT and crucial in understanding the pedagogical benefits of the WWW. Polyphony seems to be much more extensively used in technology-rich learning environments as characterised by network-based learning tools.

Towards an Ethnographic Approach to Learning and to Teaching

The research project this presentation is based on aims at upgrading the students' metacognitive level of awareness of their computer literacy. At the same time, we encourage students to adopt a kind of scientific approach to their learning processes and, on the other hand, to teaching by underlining the importance of scientific thinking even at the school level. Cognitive development is culturally-rooted and inseparable from the tools of mediation. We argue that a new kind of learning culture is about to be born. We cherish the idea of having an ethnographic approach as a learning method when using the WWW, for instance. Ethnographic approach is typical of exploring foreign cultures and it offers a model of inquiry that can be applied to classroom situations, especially when supported by computer-mediated communication and specifically designed dialogic knowledge management environments.

The research project initiated at the Media Education Centre of the University of Helsinki started with a pilot study made by Marja Mononen-Aaltonen in 1996 with an aim to know more about the learning environment as described by the students themselves. Second, we wanted to make the junior high school students more cognisant of their own learning environment. We emphasise the students' role as intelligent agents in the learning process. Therefore, they will be actively involved in our research project, which will now focus on building a dialogic learning environment on the Web by using different types of network-based learning groupware. We will look for ways of encouraging the students' intentional learning in classroom situations by making them aware of (i) their potential as researchers in the learning process, (ii) the potential of the tools the learning environment provides them for intentional learning, (iii) the potential of the mediational tools in shaping thought and communication, and (iv) the role of dialogue in learning, and in the modern networked world. Generally speaking, the emphasis has been so far mostly on the technology and on how to acquire skills in using ICT, rather than on communication. Our research aims to balance the situation by seeing the technology as a means of communication, as a modern type of mediation between human beings capable of using modern technology.

Reference

[Tella 1997] Tella, S. (1997). Multidimensionality in Media Education Tools. In Tella, S. (Ed.) Media in Today's Education. Proceedings of a Subject-Didactic Symposium in Helsinki on Feb. 14, 1997. Department of Teacher Education. University of Helsinki. Research Report. (In press.)

Distributed Web Authoring

Harold Thimbleby
Computing Science
Middlesex University
London, UK

Introduction

Authoring hypermedia documents is hard, because they are large, integrate many media, and have hypertext links and associated scripts or applets. The various media have to be kept in track of each other, creating a combinatorial explosion of version control problems; and unlike conventional media, the various components that should remain consistent need not, and sometimes cannot, all be visible simultaneously. When a hypermedia document is authored, every plan of the author can be represented in more than one place (for example, to be elaborated at the other end of a link), and each alternative development of a thought multiplies the size of the authoring problem. Very few links soon create more potential developments than can be maintained by an unaided author. Revision and maintenance of hypermedia documents further requires an author to work locally in a structure that they may not be fully conversant with. In short, hypermedia authoring is impossible do well without automated support. To provide quality control, tools have to be used.

Unfortunately, current tools have several limitations (in varying degrees) with respect to quality control:

1. They concentrate management in one person, who soon becomes a bottleneck for maintenance.
2. Or, they emphasise the appearance of pages (e.g., providing sophisticated WYSIWYG editing). This encourages diversity in stylistic design. Page editors do not scale up to handling more than a few pages.
3. Or, they use database techniques (which can guarantee consistent design and timely revision of material), but they make the design of individual pages harder, if not sterile and unrelated to page content. Database approaches typically concentrate design issues, and thereby make page designers bottlenecks.
4. Some authoring environments can visualise the site structure (which is fine for small sites, but only gives impressions of large sites), but they rarely provide any useful properties or analysis, for example that links are symmetric.

What is required is a distributed database that has a WYSIWYG user interface, and which does not centralise structural decisions, allowing distributed authors to control their parts of the structure. This would enable groups of authors to share the authoring burden, yet use database techniques to do quality control, for instance to provide a consistent navigation structure.

This paper describes a scheme that supports distributed web authoring. It allows many authors to write single pages or even large sites and unite them into a coherent site. Authors may use and share page layout designs, and these can be applied consistently. Currently the prototype system supports the design of static documents (composed of pages as sophisticated as any available HTML editors permit), but this is not a conceptual limitation of the approach. The purpose of this paper is to describe how the system works, specifically to show how powerful a simple scheme can be for creating well-organised sites out of distributed authoring contributions. We also show that the idea is productive as a tool design concept and lends itself to many extensions.

A scheme for distributed authoring

Our system is implemented as a Java program. It is run rather like a compiler, compiling 'source files' (original authored pages) anywhere in the world to 'object files,' which are given a consistent style and linkage by the compiler. We will refer to the set of object pages as a 'site'; typically a site will conceptually be a single document, in the conventional sense of having a coherent structure and message, but this is not required.

The compiler *could* be distributed, and the object files could be generated on demand. These, of course, are superficial design alternatives, and we will not discuss them further here. For clarity in this paper, we will refer to the person running the compiler as the 'user' and other contributors (possibly including the compiler user) writing web pages as 'authors.'

The compiler works in several phases:

1. the user specifies web pages (as in a browser). These pages are scanned by the compiler, and every page they link to are scanned. Scanning is subject to certain restrictions (discussed below) to stop the scanner building a database of the entire world!

2. the compiler checks the files for HTML conformance and checks all file references and performs other checks, such as images having alternative texts. These checks are summarised in an HTML report file -- this provides a very convenient form of summary because of links back to the sources of any errors in the original files.

3. source files can contain 'comments' that the compiler summarises for the compiler user. (This is a simple way for authors to remind other authors of outstanding bits of work, or for raising any other queries!)

4. more pages can be added at any stage. Possibly the user would choose a site 'home page' that gives kick off links to other pages so that the user does not need to collect root pages by hand.

5. the entire collection of web pages can now be viewed as a graph. Pages are coloured so that the user can see links to missing files and other problems easily.

6. the pages can now be compiled into target directories, organised for HTML, for images and so forth.

As so far described, the compiler is doing no more than collecting and checking distributed web pages. What is novel is yet to be described.

All source pages can contain directives to the compiler, and these control how the compiler constructs and organises the object pages into a coherent site.

Any text set between stars is treated as compiler directives; this allows the directives to be written conveniently using any HTML page editor. (A syntax that was SGML compliant would require flexible HTML editors, and would probably be harder to edit in a WYSIWYG editor.) The directives can specify:

- Comment text to be copied to the compiler user, as mentioned above.
- Defining variables, for example to associate icons with pages.
- Links to files that provide design templates.
- Directives that specify sequences of pages. These instruct the compiler that these files are to be linked linearly, where ever they might occur within the site. There are two forms:
    1. `*before*` and `*after*` can be followed by a link to a page (using HTML links as usual), and to specify that the current page should be before or after the other pages.
    2. `*sequence*` specifies a sequence of pages (not necessarily including the current page), and gives their before/after relation.
- Directives that specify nestings of pages. Again, there are two forms for nesting: `*in*` and `*contains*` for relative positioning to the current page, and `*nest*` for nesting arbitrary pages.

Directives that specify linkage

The directives `before/after/contains/in` are followed by HTML links. Thus, writing

```
*in* yyy
```

means 'put this page inside page xxx (given as a HTML link),' and it is also equivalent to '`*nest* yyy` ***me*** `*/nest*`,' where ***me*** is an explicit link to the current page.

The directives `before/after/contains/in` allow various forms of web structure to be specified. Thus, Sano recommends 'group' and 'hierarchy' for organising webs [Sano 1996]. Our `before/after` do groups; `contains/in` do hierarchy. We could add many other directives, but our purpose here is to explore the practicality of the scheme, rather than to immediately add many features that would disguise any fundamental limitations.

Suppose the compiler processes at least two of the following directives in any pages:

```
*sequence* a c */sequence*
*sequence* b c */sequence*
*sequence* a b */sequence*
```

Then a solution to these constraints is the order *a, b, c* (if only `*sequence* a c */sequence*` and `*sequence* b c */sequence*` were processed, the compiler could chose the alternative solution *b, a, c,* since the sequencing of *a* and *b* has not been specified by an author; presumably the authors do not mind if they do not say so). There are many ways of achieving the same result. The same solution would obtain from compiling page *b* if it contained:

```
*after* a
*before* c
```

The same sequence could also have been specified explicitly as `*sequence* a b c */sequence*`, and this could have been placed in any page: *a, b,* or *c* or anywhere convenient to any author wishing to impose that sequence of pages on the site.

If another author (using another page) requires that

```
*sequence* x c */sequence*
*sequence* b x */sequence*
```

then the compiler extends the solution to the order *a, b, x, c.* This satisfies all authors' sequencing requirements.

A similar process solves the nesting constraints. Finally, the compiler modifies the structure so that a `*contains*` link points to the earliest of the before-after pages any of which was explicitly contained.

As currently defined, sequences and nesting allow any ordered tree to be specified: `before/after` specifies the in-order relation, and `contains/in` specifies the parent/child relation. No page need specify complete sequences or nestings; the compiler solves multidimensional constraints to find a structure that satisfies the distributed orderings. As a special case, the entire structure could be specified by a single file, perhaps one on the user's server. That structure would specify remote files for the compiler to collect. However -- and this is one of the most important advantages of our approach - when authors wish to develop their pages, they do not need to go back to any central structure specification. If an author wants to link their page to another inside it, they could write `*contains* y` anywhere in the page. If they wanted to put *that* page inside a page that the rest of the site should refer to 'first,' it would be sufficient to write `*in* z`. This would require the compiler to place the file inside some page *z*. The compiler makes the 'tightest' linkage that satisfies the constraints: in this case it would lead to *z* being that author's 'top' page, and for the original page to be inside that.

An interesting consequence of the approach is that an author can compile their 'local' version (not necessarily geographically local) of a site independently of the other authors. They can do any quality control of their part of a larger site independently. Moreover, they can delegate parts of *their* authoring to other people, and so on without limit. (That is, the compiler can be run in many places on different components of a site, and can compile local components of sites.) Mark Addison has suggested that the compiler should compile its directives to HTML comments and, if it also parsed comments looking for directives, it could compile object pages, so they could be treated as source pages for other documents.

The compiler displays a 'dot and arrow' graph showing the site's structure. The compiler user can add new links (by direct manipulation) between pages before asking the compiler to actually compile the site. When this is done, the compiler generates constraints in exactly the same notation as authors use in their own pages: if this output from the compiler was itself compiled with the original pages, the interactively modified structure would be recreated.

Finding the site structure

Finding the site structure from the constraints is straightforward, though there are some subtleties - mostly in deciding what sort of structure one wants for a Web site in any case! For example, should cycles be permitted? In our case, we decided that cycles were inappropriate. Further, we wanted to be able to create a linear version of a site, rather like a conventional print document, and therefore required a structure that could be related to a linear document intuitively. Given these two requirements, the goal is to find an ordered tree that satisfies the constraints. A linear document is then simply a preorder walk of the tree.

Further assumptions must still be made. For example, does `*sequence* a b */sequence*` require `a` to precede `b` in the preorder, or, more specifically, for `a` and `b` to be children of the same parent page? We decided on the latter.

These decisions still leave several ambiguous cases, such as:

- If pages have no structural constraints at all, where should they go? We decided they should be made children of the root of the tree, which is (almost certainly) the 'home page' of the site; that is, they are treated as if a `*in* home` had been processed.
- If `a` and `b` are both nested in `c`, but have no order specified, the compiler has to choose one, or - perhaps - choose that `a` is nested in `b` or *vice versa*. We decided they should be treated as if a `*sequence* a b */sequence*` had been processed.

Whenever processing proceeds 'as if' some constraint had been processed, a copy of the constraint is saved to a file that can be reprocessed in the future, to ensure that the precisely same structure is preserved.

Extending the concept

A system that supports distributed web authoring has to balance a trade-off between three criteria:

- A scheme that distributed authors can use effectively. The current approach lets authors 'attach' their components of a larger document very easily.
- A scheme that has structural and design constraints (in the current case, topological sorting) that can be solved effectively and with minimal ambiguity. Also, we require error messages (e.g., when authors' constraints are incompatible) to be clear.
- A scheme that admits a good design of web pages to be generated. For example, there isn't an obvious way to design for a general graph; conversely a star or cycle is easy to design for, but has limited practical use.

We believe that the prototype balances these issues well, but leaves enough scope for useful future research! In particular, the prototype makes very few commitments to page representations, and is therefore a versatile tool. In fact the scheme described here is distributed and structurally more flexible version of an earlier system which used a centralised database [Thimbleby 1997].

Kenneth Arrow's Impossibility Theorem showed that there is a natural set of criteria for social ranking that is inconsistent [MacKay 1980]. This paradoxical result shows that there is no consensual choice algorithm for a home page that all authors would agree on (in the sense of the Theorem). Likewise, there is no consensual way of agreeing on other pages. It follows that *any* scheme, including the one proposed in this paper, is inadequate for general purposes, since no scheme can satisfy all reasonable requirements for web design. The Theorem allows for *particular* schemes for particular purposes (e.g., only one author, or all authors agree - or are required - to use a set structure), which is how most co-ordinated web sites are currently generated. Taking Arrow's Theorem more positively, it follows that any 'general' scheme such as the one proposed has an infinite number of extensions. We discuss just a few below.

Perhaps the most obvious development would be to move away from the conventional compiler model, and instead generate pages on demand. Although this would make the generated web sites more 'trendy' it would actually add nothing to the theoretical generality of the scheme. Arguably it would reduce quality control: at present, each version of the web site is generated by a deliberate and planned act of a single user - if sites were continually updated, it would be possible for authors to lose track of their versions.

As presently conceived, the compiler has to be run 'often enough.' This is not a satisfactory solution as the number of authors grows. There are many alternative arrangements, such as the compiler regularly visiting authors' sites, authors notifying the compiler (e.g., by email). It might seem ideal to completely automate compilation or to permit it to be run incrementally from anywhere. Yet large scale authoring is a collaborative activity, and it may be wiser for the compiler to provide more interactive support for communication *between* authors. For example, it would be easy to make the compiler provide direct support for user-author or author-author email, as well as compiler-author communication (e.g., for telling authors about problems with their own pages). The compiler might also track requests from one author to another (or from one author to the same author!) to undertake some writing, and (in many cases) it would know when such commitments were discharged. Fortunately these issues are orthogonal to the structural and design issues that have been solved by the present scheme.

The current compiler solves constraints sufficient to simulate any conventional print document: with ordered sections, and arbitrary nesting of sections within sections. Compared with current practice, this is passé and real web sites should be much more interesting! Though it may be useful for some applications to simulate paper documents, this is by no means the limit of a compiler approach. For example, de Bono has suggested that thinking can be usefully organised using 'six thinking hats' [de Bono 1991]. Each of the six hats is a particular colour; white is the colour associated with information, black with making judgements, and so on. Our approach could handle this, merely by extending the two sorts of relations (before/after and contains/in) to six. In this case, it may or may not be important for the reader of sites to actually see the colours. However, it would be fun to colour object pages, and have their shades change as the reader browsed. The direction of colour changes on each page could give the user a good feel for the local site structure, and what sorts of pages could be anticipated in various directions. de Bono suggests that thinking hats would help authors more than readers, and Thimbleby discusses a system that explores the advantages for hypertext authors [Thimbleby 1994].

An author may refer to a page that does not yet exist. This is not so much an error as a natural consequence of authors undertaking the huge task of writing large web sites; inevitably they will lose track of some pages that were supposed to be finished, or even created, but weren't. At present, we compile such a 'page' using a null-page design template and warn the user of the compiler - not the author of the page. An alternative would be for the compiler to ignore null pages; furthermore, an *actual* page - one an author has started to write but hasn't released for public use - could declare itself to be 'under construction' and then it should be treated like a null page. This would mean that readers of a compiled web site never saw pages that were not ready for use. More precisely, null pages can occur in two places: as leaves of a site (in which case they can safely be ignored) or as central nodes that are required to 'carry' navbars. In the latter case, the null pages either should be elided with adjacent non-null pages (so that the navigational information is not lost) or they should be compiled with a suitable design style, just to create signposts for the site.

The compiler could easily produce public and internal versions of a site. For example, an author might want to wait until they could see a draft page *in situ* before releasing it for the public site. Once such a mechanism works - being able to include or exclude pages from the site -- the criterion could be broadened: 'under construction' is not the only sort of reason for hiding a page. One might want to create 'executive summaries,' for instance, that only show the top level pages of a site; or one might want to create sites with public information and with proprietary information; and so on.

As well as being (or not being) under construction (i.e., ignored or included), pages can have very many other properties that can be used usefully by a compiler. Additional properties include:

- Whether a page should be flagged as 'new,' and if so, a date for "what's new" collections, and a text for the "what's new" summaries.
- A URL to obtain the text of a page elsewhere, for example, if the author does not have edit permission on it, but still wishes to include it in the site within the same organisational style.
- Whether a page should expire, and if so an expiry date, so the page becomes hidden in the future.
- A reminder text, that the compiler would use to track of authors' comments to themselves or to each other. For example, an author writes a reminder in a page before they take a holiday. On return from the break, they will know roughly what they were thinking before they left. See [Thimbleby 1997] for more details of reminders and their value to authors.

An author can refer to any pages using the standard `href` and other HTML references. Using a compiler, it is possible to extend the semantics of references considerably. An author may want to insert a reference to a file that contains certain text that is known to be on a particular server. A compiler can easily resolve such references.

Knowing the properties of pages, the compiler could modify links to pages in navbars or links written explicitly by authors or by other means so that they had consistent and appropriate style. If a page is flagged as 'new' then some (possibly all) links to it could have an associated new icon. If de Bono's hat colours were used on pages, then the links to them could be coloured appropriately.

Many resources used in a large document require significant work from the author. A simple example is the provision of alternative (`alt`) text for images or a width and height for them. A compiler can easily provide a built-in database of such details, and ensure that they are used consistently (unless locally overridden, in which case it might provide a suitable diagnostic).

Other features of the compiler (unrelated to structure)

The compiler provides other features to support distributed web authoring, and for completeness we mention them here.

The compiler creates a graphical image of a site, which the user can edit and manipulate in many ways (mentioned briefly above in relation to editing site structure). They can view ranked embeddings, and see structure in a site that is very helpful in evaluating its design. Such views might also be useful for readers of a page, as an active map, so that they better understand where they are within a site and where they can go. We ought to extend the compiler so that it can generate image maps for use in navigation. The compiler can also draw many statistics charts, such as byte count (i.e., page download time, including images), page eccentricity and other graph theoretic measures.

Variables may be defined in any files, and their values obtained by a simple inheritance mechanism (using the tree structure of the site). Thus it is easy to refer, say, to the icon (i.e., an author-defined variable name) of any page, and in particular, it is easy to refer to the icon of the next, previous and up and down pages - using a simple syntax that allows the values of variables to be obtained from other files.

The value of a variable defined in any file is not sufficient. It is also necessary to generate links to the appropriate files. The compiler provides a mechanism for retrieving the file name where a variable is defined, and using it as a HTML link.

Code	**Meaning**
`*get* var`	Value of *var*.
`*ref* var text */ref*`	Make *text* a hot link to where *var*is defined.
`*refget* var`	Abbreviation for `*ref* var*get* var*/ref*`.

There are a range of *ad hoc* features that ease writing web pages. For example, writing `*symmetric*` before any anchor converts it to a symmetric link: given a href/name link in either direction, the compiler generates href/name links in the other direction, hence making the link symmetric.

There are a wide range of built-in variables that provide useful information, such as the date. However, to avoid the compiler accreting a wide range of arbitrary features, Java objects (applets) can be loaded and run. The compiler passes the object parameters from the source file and inserts the stream output of the object into the compiled HTML. Java provides a clean interface so that authors can do almost arbitrary things to suit their own needs, to share libraries of features, and do so without the compiler getting more complex or harder to maintain.

Conclusions

We have described a very simple but powerful scheme for organising distributed web authoring. Individual authors can write pages or create design elements. The structure of the overall site can be specified in one place, or it can be distributed. Pages can specify where they wish to be within the overall structure, and the compiler solves constraints to find the most compact structure that satisfies the authors' requirements. Each page inherits a design template from the structure, and this ensures a consistent web site design, yet allows considerable flexibility. Design templates can place navigation menubars and other features so that users can browse the structure of a site conveniently.

The current scheme is a powerful approach on which to build more sophisticated authoring tools. We argued that the number of extensions was endless in principle, and we gave a few examples of extensions to explore the potential for development.

Acknowledgements. This project is funded by EPSRC under grant number GR/K79376. The author is grateful for input to the ideas expressed in this paper from George Colouris, Gil Marsden and A Min Tjoa.

References

[de Bono 1991] E. de Bono (1991).*I am right. You are wrong.* Penguin Books.

[MacKay 1980] A. MacKay (1980). *Arrow's Theorem: The paradox of social choice.* Yale University Press.

[Sano 1996] D. Sano (1996). *Designing large scale web sites.* John Wiley.

[Thimbleby 1994] H. Thimbleby (1994). "Designing user interfaces for problem solving, with application to hypertext and creative writing," *AI & Society*, **8**, 29-44.

[Thimbleby 1997] H. Thimbleby (1997). "Gentler: A tool for systematic web authoring," *International Journal of Human-Computer Studies*, S. Buckingham Shum and C. McKnight, eds., **47**, 139-168.

Web-based Virtual Learning Environments: Experiences and Futures

Morten Wagner, Jørgen Bøegh and Allan Meng Krebs,
DELTA Danish Electronics, Lights & Acoustics – {mw,jb,amk}@delta.dk

The web provides a number of advantages for distance learning. This short paper describes and discusses the evolution and current design of the learning environment developed by DELTA Danish, Electronics, Lights & Acoustics and also identifies trends and futures of web-based learning environments in general. A more detailed description can be found in [1].

Different web-based learning environments

The web gives educators the possibility to integrate proven as well as new methods of teaching. From mail-based do-it-yourself courses to interactive audio- and video conferencing – the web has the potential to integrate all. In web-based learning environments, the classroom needs not be further away than the nearest Internet-connected PC. Many different designs and implementations of web-based learning environments have already appeared. Amongst those technologies we have seen in use are;

- Usenet-like, asynchronous discussion groups
- eMail communication
- publication of material on the web
- shared pools of documents
- video- and audio conferencing
- moo's and muds

These different technologies are combined in different forms to represent more or less explicit learning environments: sets of tools, that in combination define the students' opportunites and modes for learning. We have had our focus on designing and implementing a coherent and user-friendly environment, in which the user does not consciously have to recognize the different modes of teaching and learning, but can utilize already present learning-skills. Our design has evolved a lot since the start; trials and questionnaires have formed a based for constant improvement and changes.

Traditional computer-based training v. web-based learning

In traditional computer-based training (CBT), instructional programs are most often being designed specifically towards some area. When a CBT-course has been implemented and distributed, it cannot be changed or updated, no social contact is integrated (and often not necessary) to complete the course. With web-based learning, both ordinary CBT can be deployed *as well* as centralized control and interactivity with fellow students.

Synchronous and Asynchronous modes of learning in the design of web-based learning environments

As part of our goal to satisfy different modes of teaching and learning, we have seamlessly integrated both asynchronous and synchronous modes in our environment. The integrated asynchronous modes include using interactive self-learning material, participation in usenet-like conferences, and sharing of documents. The synchronous modes supported includes on-line teaching and presentation, group-work facilities, etc.

A challenge in designing an integrated environment for learning, is the natural combination of these different modes of learning and teaching. Most often, many different tools and skills are needed to exploit different modes, but we have emphasized a more natural integration, integrating all interaction with tools in a web-browser. As part of the web's nature, continuing experimentation and evolvement of the interface are easy and straightforward – leaving the interpretation of result in field-trials as one of the big challenges.

From the field: Experiences with different modes of learning

Our virtual learning environment is built up around the metaphore of a virtual campus. From the campus hallway, the students can enter different rooms with different functions. Online presentations take place in the classroom, online cooperative work takes place in the grouprooms, and we have billboards and posters with discussions and information. Every room and billboard have it's different functions and support different modes of teaching and learning. Below, some important points from our latest trial are summarized;

- **The classroom**: In the classroom, online lectures and presentations are given with audio- and video conferencing. The lecturer can show prepared slides on a shared, virtual "slide-projector".
  Experiences: Students are generally positive about this form of teaching. There are, nevertheless, some differences in the students' preparedness for asking questions. Some types of students never ask questions in this environment, others are heavily engaged in discussion, trying out the boundaries of the new media. The "communication distance" between student and teacher is perceived as smaller than in the auditorium but still greater than in the classroom. Another result is that distance teaching and presentation demand quite a few new skills from the teacher; students must be coached into using the media, the presentation material must be more engaging, etc.
- **The grouprooms**: The grouprooms are a collection of tools that allow students to cooperate on solving a specific exercise. This can include educational material, needed for the solution.
  Experiences: Generally, our students have some difficulties in initiating and performing efficient groupwork. Some students seem somewhat alienated towards the environment, and make no actual attempts to engage in groupwork.
- **Posters and billboards:** The billboards act as asynchronous means of discussion and information exchange.
  Experiences: These tools are effective and easy to use and accept. It is necessary, though, initially to engage the students, for example by giving them assignments that include using the billboards and posters.
- **The Tearoom:** The tearoom is a place for social (and chance-) meetings. This has not been sufficiently tested yet.
- **The study:** Here teaching material, other related material and the students own papers are stored. All selfstudy material is made highly interactive, with self-tests, indexes etc.
  Experiences: Though it is expensive to construct good selfstudy material, the learning effect is good and the students are satisfied.

Generally, online groupwork, teaching and presentations are motivating and good tools in web-based education. Surprisingly, it is very diffucult for people initially to engage in groupwork; it is seen that even computer-literate users have difficulties in mustering initiative and collaboration skills. Students often feel alone in this environment, and thus video representations of fellow students are important. In some cases, though, video can be a distraction; especially in some types of lectures and presentations.

In online groupwork and communication everybody is more equal; it is only possible to dominate a group by verbal behaviour – not by other means.

The self-paced, asynchronous parts of the web-based environment are much easier for users to exploit effectively. Using the interactive, synchronous parts of the environment is an area for further study and experimentation. Indications are also that the CSCW tools are not quite good enough, yet. We believe that more initial coaching and "safe" environments will ease the transition from real to virtual classrooms.

Trends and directions for web-based learning

Now, more and more online conference tools are emerging. These meta-tools are including more and more tools like shared whiteboards and shared applications. In our case, we have developed our own shared whiteboards and editors, our own presentation tool etc. This allows us to control the integration with the web-based, asynchronous material, and to create a homogenous environment. As the "big companies" continue their development effort in this area, it will probably be possible more precisely to define interface and integration with the web and thereby have more specialized (and easier) environments for the students to use.

In general, the web allows us not only to generate web-based learning material, group-work, etc., but can provide the students with actual environments, with informal chats, "chance meetings" and both controlled and uncontrolled exchange of knowledge. Discussions can be carried on in online and offline forums and the learning is not confined to one or two specific modes.

Acknowledgement

The work described in this paper was done as part of the ACTS project Prospect. Prospect is partly funded by the European Commission.

References

[1] Tele-Educational Services in a Future Open Service Market, *Mark Riordan & Vincent Wade, Trinity College, Dublin, Allan Meng Krebs, Jørgen Bøegh & Morten Wagner, DELTA Danish Electronics, Lights & Acoustics.* Proceedings of ED-MEDIA, ED-TELECOM 1997

Conceptualizing a Tri-Modal Model for Web-Based Distance Education

Joel Wilkinson

Project Manager, Media Learning Systems

Community Rehabilitation Studies

The University of Calgary, Calgary, Alberta

joelwilkinson@pobox.com

http://rehab.educ.ucalgary.ca

Jonathan L. Ross

Program of Educational Technology

Graduate Division of Educational Research

The University of Calgary, Calgary, Alberta

jross@acs.ucalgary.ca

http://www.ucalgary.ca/~jross

Abstract: The Internet offers tremendous potential as a medium for delivering educational course material. This paper will discuss the development and implementation of a multidimensional model for program delivery via the World Wide Web. Included is a description of the evolution of three distinct delivery modalities that take into consideration course content, student needs, and the mandate of our community-based program.

Introduction

Using the Internet as a vehicle for distance education has become increasingly popular in recent years [James & Gardner 1995; Ibrahim & Franklin 1995; Bigelow 1996]. Post-secondary institutions, in particular, are harnessing the capabilities inherent in the Internet, creating courses, and in some cases complete degrees, for delivery on the World Wide Web [Dimitroyannis, 1994] .The Web has a number of inherent advantages as an instructional medium. Most significantly, course materials can be viewed irrespective of geographical location and time of the day (see Goldberg, Salari & Swoboda, 1996). Furthermore, a large number of students can be served simultaneously, thereby reducing the costs to academic institutions [Goldberg 1996].

The Bachelor of Community Rehabilitation Studies "Community of Learners" model encourages students, located in various geographical regions throughout in Alberta, to collaborate with each other. The University facilitates these collaborative sessions, helping students become actively involved in their learning. Supported by a provincial governmentn ACCESS grant, our program is a partnership with six other community colleges in the province.

Conceptualizing Our Model

We have managed to create nine courses delivered in part, or in whole, via the Internet. Our model of program delivery is simple, yet effective. The process of developing and delivering on-line courses has evolved intocan be conducted in three distinct modalities: three fashions. Institutions can offer distributed educational opportunities using:

Full Internet Delivery;

Internet Enhanced Delivery; and,

Internet Supported Delivery.

Full Internet Delivery (Currently utilized by three Health Foundations courses--EDIS 551.58,551.59,551.60)

This modality lends itself to the inquiry-based learning model where course instructors become facilitators of students growth and development as learners, researchers and practitioners. These courses use the Internet as the primary vehicle for delivering all material on-line. Multimedia such as video case studies, graphic and sound files provide students with a number of ways of learning, both process, and content.

One of the most common criticisms of on-line learning is the lack of collaborative and social opportunities for learners [Goldberg, 1996]. To help address this issue our full Internet delivered courses allow students to engage in both synchronous and asynchronous text-based communication. Using Java technology, students can conduct on-line seminars in private or semi-private environments. Newsgroups and bulletin board systems provide asynchronous communication interactive opportunities.

All assignments are administered and evaluated on-line. Students, working with case studies, are invited to complete text forms that are compillied and sent to the course instructor for evaluation purposes. Instructors then follow-up with email communication and comments thereafter. Future directions for these courses include establishing a database for student assignments and instructor comments to allow for seamless access to all course materials, assignments and evaluations.

Internet Enhanced Delivery (Currently used by four Educational Psychology courses--EDPS 415,425,573,581)

Using the Internet enhanced model of program delivery, content is presented via video conferencing, the Internet, and weekend workshops. Content such as the course outline, assignments, related links, and email lists are posted on the Web as reference tools. Synchronous and asynchronous chatting capabilities are also provided on the Internet to stimulate collaborative group work and critical discussion. However, tThe bulk of the content is delivered through video conferencing and occasional weekend workshops. This balanced approach to presenting content allows students to benefit from social/collaborative opportunities in-person workshops provide while having the Internet serve as as a course-enhancing resource tool. As a result, students from various geographical locations are able to work together on particular projects.a supplementary resource tool.

Assignments can be completed on-line. Students, using synchronous and asynchronous communication methods are able to discuss issues and problems case studies raise. Assignments may then be delivered via email to the course instructor for marking.

Internet Supported Delivery (Currently used by two Educational Psychology courses-- EDPS 589.02, 475)

Although the Internet is still used to provide students with supplementary course information, content is delivered primarily through in-person workshops and classes. Using this particular modalitymodel, the Web becomes a resource area for students where information is presented such as the course outline, reference links, instructors email, and related course Newsgroups. The course homepage links the instructor and the students between classes. Important information relating to the course may be posted on the Web site for students to read, or relayed to students using email.

Implementation

Successful implementation of the tri-modal models requires a team approach to the design, delivery and evaluation of our Internet-based courses. The Media Learning Systems group

(MLS), responsible for the implementation of the various distance education projects, is comprised of experts in the following areas:

Project Manager (coordinates entire process)

Course Content Expert - Instructor

Instructional Designer (manages the process)

Web Administrator (technical design/support)

Graphic Designer/Multimedia Developer

HTML Programmer (web content mark-up)

Initial meetings are held where instructors needs are balanced with the current technology capabilities to ensure that course and program goals are met. It is imperative that instructors have an understanding of current Web technologies when planning for, and administering, on-line courses. Hence, time is spent on the training and development of staff so that they can make well-informed decisions regarding the use of technology.

Once the instructors course objectives have been balanced with the current technological capabilities, regular development team meetings ensue to construct action plans for developing the course shell.

The next step is to have individual sessions with instructional designer and members of development team for more specific aspects of the course requirements. The instructor and instructional designer are responsible for the development and modification of course content. The development teams responsibilities include the design the course website template; insertion of course content; instructor training in utilizing the technology features; and evaluation, modification and revisions of the course website as directed by the instructor.

Future Objectives

Providing students with an on-line orientation would serve as a beneficial training tool. We are currently working on a site tour that will orientate students prior to commencing the course. Instructions for using the communication tools, navigating and completing forms are some of the sessions that will be built into this area.

We found that there were some problems with students having compatible systems and software. This affected students access to course materials and caused unnecessary frustration. Thus, it is important to *ensure students have access to appropriate technology (hardware/software) by publicizing system requirements prior to the courses commencement.* In this way all students are aware of what is needed to complete the course.

At the same time, however, we do not believe it is fair to expect students to purchase expensive and complicated equipment and/or software required to access our materials. While it is imperative that our courses and delivery methods reflect the latest technological advances, we also must *ensure that our courses are user-friendly and do not burden students with unfair technical requirements.*

We are working on ways to help instructors become more responsible for directly creating and maintaining their course web sites. Currently, we are developing *a user-friendly web-editor to allow instructors to post and edit content from their desktop.* For this system to work, however, instructional design support must be provided to course facilitators. An effective lecturer may not necessarily be an effective Web developer. Training and support are required to ensure the transition process is successful.

Conclusion

While the Web continues to develop in its ability to efficiently and effectively delivery materials to students studying at a post-secondary level, there remains the need for educators to evaluate critically evaluate the way they use the medium. It is hoped that our tri-modal model can give developers informed choices in the way they use capitalize on the potential of the Internet to deliver educationalcourse material.

References

[Bigelow 1996] Bigelow, J.D. (1996). Developing an internet section of a management course: Transporting learning premises across media. Proceedings of the 1996 World Conference of the Web Society, 13-20.

[Dimitroyannis 1994] Dimitroyannis, D. (1994). Virtual classroom: A case study. (On-line) http://www1.cern.ch/PapersWWW94/ddimitri.ps

[Goldberg, Salari & Swoboda 1996] Goldberg, M.W., Salari, S, & Swoboda, P. (1996). World wide web course tool: An environment for building WWW-based courses. Computer Networks and ISDN Systems, 28, 313-317.

[Goldberg 1996] Goldberg, M. W. (1996). CALOS: An experiment with computer-aided learning for operating systems. Proceedings of the 1996 SIGCSE Technical Symposium.

[Ibrahim & Franklin 1995] Ibrahim, B., & Franklin, S. (1995). Advanced educational uses of the world-wide web. Computer Networks and ISDN Systems, 27, 871-877.

[James & Gardner 1995] James, W.B., & Gardner, D.L. (1995). Learning styles: Implications for distance learning. New Directions for Adult and Continuing Education, 67, 19-31.

Our web site can be viewed at: http://rehab.educ.ucalgary.ca

*j*Input - A Java applet for CJK keyboard input method

Leong Kok Yong (kokyong@irdu.nus.sg)
Liu Hai (liuhai@irdu.nus.sg)
Oliver Wu Peng (oliver@irdu.nus.sg)
Internet Research and Development Unit
National University of Singapore
10 Kent Ridge Crescent, Singapore 119260
URL: http://www.irdu.nus.sg/multilingual/jinput/

Introduction

Web internationalization (i18n) has made much headway on the Internet scene since the Web inception. New features in HTML 4.0 like language tags will make the Web more international. Support for HTTP content language negotiation are being built into both client browser and web server. Towards these effort, Netscape and Microsoft are building more support in their web browser for viewing various language encodings. At the last count, Alis Technologies's Tango browser has support for display of the widest range of language encodings, including right-to-left languages and keyboard input methods; but unfortunately it is available for Windows only.

On the other hand, true i18n keyboard input methods on many platform is still limited, although commercial application abounds. Windows users would purchase and install third-party helper applications which allow them to type i18n text in any Windows application, including the web browser. Macintosh users would need to purchase the Apple Language Kits to get i18n input methods. UNIX do not have any such system add-ons. The next upcoming WindowsNT 5.0 though has promised multiple IMEs on one single system but users will have to wait.

Background

At the present moment, users will use helper application to enable them to input characters that is not native to their operating system platform. For example, a US-English Windows user wishes to submit some Chinese, Japanese or Korean (CJK) phrases as keywords to a web search engine. He needs to input these characters into the text field of the Web form. He may choose to download and install a helper application (like WinMass, Unionway or NJStar) to enable him to input CJK characters. So unless he is using a native Chinese/Japanese/Korean Windows operating system, he may need to download and install one of these application (typically around the size of few hundered Kbytes to few Mbytes). A Macintosh or UNIX users may not be so fortunate as such applications are not readily available. To overcome this, we developed *jInput*.

jInput

We use Java for the user-interface component that allows the user to input their keywords. As a first prototype, we developed a Java applet that enables web browsers to accept Chinese characters input in GB encoding. With Netscape's LiveConnect technology in their Navigator and Communicator browser (which allows Javascript to call Java methods), the applet's GB encoding content can be passed back to a form variable before all variables and their values are submitted via CGI.

Why use Java?

Java is cross-platform! By using Java, a platform independent language, our applet can run on Windows, Mac and UNIX platform which supports a Java virtual machine. For instance, Netscape browser offers you the ability to run Java on Windows, Mac and UNIX. In addition, we noticed the lack of keyboard input method support in Java. Even with the most recent version, JDK 1.1.x, the furthest reach into the internationalization features is the support for font display of i18n text.

Implementation

Current implementation allows a user to input CJK text with the applet. This includes (1)Chinese GB using the PinYin and CangJie methods, (2)Chinese Big5 with PinYin, CangJie and Simplex methods, (3)Japanese JIS with RomanKana and Tcode methods, (4)Korea KSC with Hanja and Hangul methods. These input methods are based on the Chinese Xterm - cxterm. For a demonstration copy of the Java applet, please see http://www.irdu.nus.sg/multilingual/jinput/.

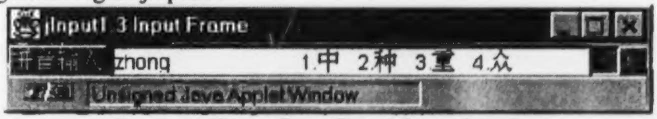


Figure 1: jInput InputPanel frame/window

The applet is organized into 3 packages - *method*, *font* and *gui* packages. The *method* package provides the relevant keyboard input functionality, mapping keystrokes to the corresponding characters. The *font* package provides the corresponding bitmap glyphs for drawing each character which the user enters. Lastly, the *gui* package provides the GUI components, similar to the TextField and TextArea in the Java awt package but enhanced with the CJK input method and fonts, e.g. we have a single line text field component - '*EditingField*' which is the CJK enhanced version of java.awt.TextField component in JDK 1.0. By dividing the *method* and *font* components from the *gui* components, we can easily adopted different input methods or font information into the *gui* components.

Ongoing Development

To build upon our existing ideographic CJK support, we are developing the next version of our framework to support most of the world's writing system, including right-to-left (Arabic, Hebrew), Pan-European (French, German, Greek, Cyrillic, etc.), Indian (Tamil., Devanagari, etc.), Thai, etc. To achieve this, this next release will make use of Unicode and the JDK1.1 internationalization features.

Our development work so far is carried out in Java JDK 1.0 since most browsers currently only support Java 1.0 release (as of this writing Netscape has released a preview release of a JDK 1.1 patch for it's Communicator 4.02 browser for Windows). With the release of Java *JDK 1.1*, support for internationalization has improved, in particular the capability which allow a Java application to output text using non-system fonts is very useful in our context. The largest component in our applet are the font information - bitmap glyph for each of the CJK characters. Since more and more users are expected to install fonts for specific language they wish to view in their browser, the next version of the '*jinput*' Java applet could read font information from the fonts installed on the client computer instead of downloading these huge amount of font information.

The drawback of the applet is that it is useful only when running on Netscape 3.x / 4.x which supports LiveConnect (Javascript to Java intercommunication). To extend its coverage, one possibility of using VBScript to interface to the Java applet for Microsoft Internet Explorer is being studied.

Because of its sheer size (average of 200kb ~ 400kb), its use as an applet is not very optimal. To make the system even more useful, we are porting it from an applet to an application. This will hopefully be released very soon as a Netscape Composer plugin to allow users to input CJK text when publishing HTML pages using Composer.

To make it easy for Java developers to reuse these Java input methods components, we intend to make them available as JavaBeans.

Conclusion

We have illustrated that it is possible to write a user interface component for i18n keyboard input method that runs on multi-platform using Java. We hope that with these Java classes as the basic framework, more application can inherit these keyboard input methods using JavaBeans.

Acknowledgment

We thank Dr Tan Tin Wee who first put forward the concept of developing a Java applet with built-in keyboard input methods and font display capability to us. We appreciate Aaron Aw for his effort in helping with the Java stuff and Netscape Live-Connect technology.

BEST COPY AVAILABLE

References

[REAZ HOQUE 1997] *Java, Javascript And Plug-In Interaction Using Client-Side Liveconnect,* http://developer.netscape.com/library/technote/javascript/liveconnect/liveconnect_rh.html
[Cxterm] *Cxterm (Chinese Xterm) source code,* ftp://ftp.ifcss.org/pub/software/x-win/cxterm/

Visualization Tool for Collaborative Web Browsing

Guillermo S. Zeballos and Marian G. Williams
Computer Science Dept., University of Massachusetts Lowell
One University Avenue, Lowell, MA 01854 USA
{gzeballo,williams}@cs.uml.edu

Introduction

This paper describes work in progress to develop visualization tools to support *collaborative Web browsing*. Collaborative Web browsing is the act of information gathering on the *World Wide Web* (WWW) by two or more individuals working together with a common goal. Their activity may involve locating information, giving a guided tour, or engaging in recreational browsing. For all such activities, users need to coordinate their efforts. While some collaborative Web browsers and tools for collaborative Web browsing have been developed, there is a need to develop tools for visualizing the browsing activities of the participants. In the physical world, participants in team scouting make use of maps to note landmarks and inform them of their relative positions. Since maps of Web space are impractical, we envision visualizations that show participants where they have been, rather than all the places where they could go.

Recent studies of browsing patterns show that the current stack histories of most browsers do not adequately support navigation on the Web [Tausher & Greenberg 97]. One of the problems is that they do not provide information about the context in which a page was viewed [Furnas 97]. More sophisticated ways to visualize browsing histories are needed. Work along these lines has been carried out by [Ayers & Stasko 95], using n-trees to illustrate paths navigated by a single user during a Web browsing session. Also, collaborative Web browsers have been developed to allow multiple users to access Web sites simultaneously from different terminals. *GroupWeb* [Greenberg & Roseman 96] and *W4* [Gianoutsos & Grandy 96] have shown the usefulness of Web-based groupware tools such as multiple cursors, shared annotations, and shared browser controls. The browsers themselves are designed to slave multiple clients to a single site. Our goal is to support collaborative browsing where the browsers themselves are spread over multiple sites and when users are not forced to work in lock-step.

Collaborative Browsing and Visualization

Collaborative Web browsing is a useful whenever the browsing task is too large for one person to perform alone (for example, when a search query returns a large set of candidate links) or when the expertise of more than one individual is necessary for locating and/or selecting useful information (for example, when a teacher is guiding students in their browsing, when multiple users have expertise in different domains, or when an expert needs to train a novice.)

In all of these scenarios for collaborative Web browsing, two or more users working on a collaborative browsing task may spend time looking at different Web pages, but may also choose to examine the same pages synchronously when doing so seems beneficial. In order to coordinate their work, they need to be aware of the current state of each other's browsers. They also need an awareness of where in Web space each of them has browsed, including where their browsing activities have overlapped. Awareness of this kind may be provided by a visualization tool. A history tree, as described by [Ayers & Stasko 95], represents the navigational space that has been visited by a single user.

We are extending Ayers and Stasko's work for collaborative Web browsing. A collaborative history tree will enable each participant to visualize not only where he or she has browsed, but also where the other users have browsed and where the various users' browsing has overlapped. The collaborative history tree will also allow a user to quickly access a fellow browser's current page.

The collaborative history tree is being developed in Tcl/Tk [Ousterhout 94]. A preliminary version has been written to drive a Netscape[1] browser running in a UNIX environment. It uses Netscape's remote control facility, which allows for external manipulation of the browser. As new sites are visited, a tree is built. Each site visited becomes a node in the tree. The nodes themselves act as hyperlinks; clicking on a node causes the browser to jump to a new site. Since jumping makes the browser's own history unpredictable, the browser's back/forward facilities are disabled.

The preliminary version suggests that the trees should be context dependent, so that they do not necessarily record all browsing activities. Context-based history records are recommended by [Tausher & Greenberg 97]. We plan to provide context-sensitivity by creating a paging facility that will load separate trees onto the display. Collaborative history tree editing facilities will also be provided. The collaborative history tree capability will be added to a collaborative Web browser (GroupWeb [Greenberg & Roseman 96]) and it's practicality as an awareness tool for collaborative Web browsing will be evaluated.

Evaluation plan

We are interested in answering the question, can collaborative history trees enhance the effectiveness of collaborative browsing activities? We will perform a study to compare browsing efforts with and without the collaborative history trees. In both conditions, test subjects will be able to communicate orally, as if on the telephone, but will have no visual contact. The tasks performed in the study will include scavenger hunts for topical information, using the results returned by search queries, and facilitation activities, in which a more experienced user will guide a less experienced user. We will measure accuracy in accomplishing the collaborative browsing task, speed of accomplishing the task, the nature and quantity of spoken interaction, and test subjects' opinions of the collaborative history trees.

References

[Ayers & Stasko 95] Ayers, E. & Stasko, J. (1995). Using Graphic History in Browsing the World Wide Web. *Tech. Report GIT-GVU-95-12*, Graphics, Visualization, and Usability Center, Georgia Institute of Technology, May 1995.

[Furnas 97] Furnas, G. (1997) Effective View Navigation, *Proceedings of the ACM SIGCHI 1997 Conference on Human Factors in Computing Systems (Atlanta GA, March 22-27)*, ACM Press, pp. 367-374.

[Gianoutsos & Grandy 96] Gianoutsos, S. & Grandy, J. (1996), Collaborative Work with the World Wide Web: Adding CSCW support to a Web Browser. *Proceedings of Oz-CSCW'96 Brisbane, Australia, August 1996.*

[Greenberg & Roseman 96] Greenberg, S. & Roseman, M (1996) GroupWeb: A WWW Browser as Real Time Groupware. *ACM SIGCHI 1996 Conference on Human Factors in Computing Systems, Conference Companion*, ACM Press, pp. 271-272.

[Tausher & Greenberg 97] Tausher, L., & Greenberg, S.(1997) Revisitation Patterns in World Wide Web Navigation, *Conference Proceedings of ACM SIGCHI '97 Human Factors in Computing Systems (Atlanta GA, March 22-27)*, ACM Press, pp. 399-406.

[Ousterhout 94] Ousterhout, J. (1994), *TCL and the Tk Toolkit*. Reading, MA: Addison-Wesley.

[Tausher & Greenberg 97] Tausher, L., & Greenberg, S.(1997) Revisitation Patterns in World Wide Web Navigation, *Proceedings of the ACM SIGCHI 1997 Conference on Human Factors in Computing Systems (Atlanta GA, March 22-27)*, ACM Press, pp. 399-406.

---

[1] Netscape is a trademark of the Netscape Communications Corp.

Tiger: Educational Research in the Age of the Internet

Yong Zhao, Department of Counseling, Educational Psychology and Special Education, Michigan State University, USA, zhaoyo@pilot.msu.edu

Valerie L. Worthington, Department of Counseling, Educational Psychology and Special Education, Michigan State University, USA, worthi14@pilot.msu.edu

Richard E. Ferdig, Department of Counseling, Educational Psychology and Special Education, Michigan State University, USA, ferdigri@pilot.msu.edu

Presenting at national conferences represents a vital component of being an educational researcher. Researchers who present to groups of their peers have the opportunity to disseminate and discuss their original ideas, which helps to push their own thinking and subsequently the field of educational research. However, as most researchers know who have given presentations, the act of submitting proposals to conferences can be a very stressful experience. As often as not, researchers find themselves working until the last minute to perfect and photocopy the proper number of proposals; dashing to the post office or overnight delivery service to send the proposals on their way; and hoping after the fact that everything gets submitted properly. Although this process can be unbelievably stressful and inconvenient, this stress and inconvenience has come to be viewed as part and parcel of the submissions process, and hence, of what it means to conduct scholarly research and be a researcher.

However, in the age of the Internet, it is possible to simplify the proposal submissions process. E-mail, file transfer protocols, and the World Wide Web are all electronic methods of sending and receiving information that cut down on paper use, are never in danger of being lost in the mail, and can reach a destination in a matter of seconds or minutes rather than days. Importantly, in addition to these improvements in efficiency, the advent of electronic media also has the potential to challenge the beliefs of researchers about what it means to conduct and share research.

Our supposition about this challenge to existing practice is predicated upon our experiences developing and pilot testing, TIGER, a Web- and e-mail-based proposal submissions process for the American Educational Research Association (AERA). In developing the process, which enables AERA members to submit presentation proposals for the annual meeting, we anticipated that the process might serve as a catalyst for changing the way the entire organization views the conduct and dissemination of educational research. What we did not anticipate, but what manifested itself in the interaction between users and the system, was a negotiation between users and developers of the system as to the kinds of functions that the system should be able to support. Ongoing feedback from users indicated that they were willing to adapt to some aspects of the system, but that they also insisted upon adaptations on the system's part. This ongoing negotiation of the purpose of the system structures an underlying negotiation of the meaning of research. Data--in the form of on-line evaluation information, e-mail messages from users, and feedback from the ad hoc committee that was formed to investigate the possibility of on-line submission--indicates that this negotiation process is indeed taking place. We focus on the practical and symbolic significance of this negotiation process, as well as its implications for the future organizational norms, beliefs, and activities of AERA members.

SIG Discussions

Using the Internet for and by Teachers

Susan Johnson	Bill Rudolph	Janet Sharp
Mathematics Department	Mathematics Department	Curriculum and Instruction
Ankeny High School, USA	Iowa State University,USA	Iowa State University,USA
sjohnson@iastate.edu	wbrudolp@iastate.edu	pipurr@iastate.edu

This discussion session will focus on the technologically-enhanced methods of interaction and instruction among mathematics teachers taking graduate courses through distance education. This master's degree program includes activities whereby students develop a sense of community and subsequently feel an ownership and kinship with other learners and with the program, in general. The internet, e-mail, and list servers are important mediums for the activities in this integrative approach to learning at remote sites. Such activities eliminate the alienation which is frequently part of distance learning. In fact, these learners use the methods and technologies studied in the courses to develop a viable educational environment through existing technology available to their own secondary mathematics students. As a consequence of their own experiences, these teachers create an educational environment in their individual classrooms with the potential to markedly influence our present educational system.

Regarding the former situation, we will discuss the efforts of a midwestern university to provide professional development, educational opportunities to secondary mathematics teachers throughout the state of Iowa. Regarding the latter situation, we will present an action research study conducted by one of those graduate students, which reflects the motivation and experience gleaned from this university experience. She successfully implemented internet technology into a secondary school mathematics course and she will discuss the ramifications of this experience from her viewpoint and the viewpoints of her students.

Fiber optics distance education links over 395 sites in Iowa and has provided the means for learners throughout Iowa to take courses for college credit and complete degree programs while attending the site in or near their own communities. The fiber-optic connections not only permit two way real-time interactions but also allow remote site participants to communicate with others in the class or around the world using list servers, e-mail, WWW links and facsimile transmissions.

Inherent in the activity of learning via distance education and the many associated technologies is the opportunity for educators to learn how to positively influence the learning of their own students. Teacher-participants learned methods of (1) investigation, (2) reasoning, (3) communication of findings to others and (4) presenting context-based problem solutions. These four goals are consistently espoused by the National Council of Teachers of Mathematics in the *Curriculum and Evaluation Standards for School Mathematics*. Such experiences with technology change what teachers are teaching, the way they are teaching and the way their students are learning.

These experiences from the university distance education coursework led one of the participants to use the internet in her classroom. The internet provided her students the opportunity access data resources leading to (1) posing motivational mathematics problems, (2) organizing data collection, (3) developing hypotheses about interrelationships, (4) communicating with others about their findings, and (5) discussing the implications of their research. Because her students found these technology-based activities motivational and thought-provoking, we may conclude that positive changes in the educational environment due to appropriate uses of technology are not only possible but likely.

BEST COPY AVAILABLE

WEB BASED COURSE DEVELOPMENT: ADVANTAGES AND CONCERNS

Charles H. McCloy, Instructor, History/Political Science, Trident Technical College, USA, zpmccloy@trident.tec.sc.us

Nancy Simpson, Instructor, Psychology, Trident Technical College, USA, zpsimpson@trident.tec.sc.us

We have been using the WWW in teaching for over three years. The following are some of our observations and comments we have heard from our students.

Advantages:

Empowers - The World Wide Web enables students to learn and explore areas of interest within a structured framework rather than in the sometimes passive environment of the traditional classroom.

Flexibility - Allows students who work different schedules to continue their educational goals. This is also advantageous for homebound students, students with small children and those students that have a conflict in scheduling.

Additional Instruction - In our courses the instruction via E-mail has increased contact with the students. We have found that students are less inhibited , ask more questions, and make more comments than those in the classroom.

Critical Thinking - Rather than simply taking notes, it forces students to analyze information on their own.

Relevancy - Provides timely information which brings the course alive to many students who often do not completely see the relevance of the material in the textbook.

Concerns:

Computer Knowledge - Students who take the course must have a basic knowledge of computers. If an instructor is not careful the course can become a "how to" on computers to the determent of the course material.

Less Work - Some students see the course as an easy way to get three credit hours.

Disciplined - On line courses requires students that are self motivated to work on their own.

Fad - Some students take the course because it is a new idea or method of education. When the newness of the concept wears off they drop out.

Instructors - Required the ability to write well and be concise. In the traditional class the instructor can rephrase a question if it is misunderstood. In an on line course this is more difficult to do without some students becoming confused and frustrated.

Testing - Necessary to ensure the academic integrity of the program.

Cost - Due to equipment cost on line courses may not be available to lower income students unless the equipment is available on campus.